

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: March 14, 2006, 02:19:42 ; Search time 6853 Seconds

(without alignments)  
4424.052 Million cell updates/sec

Title: US-10-620-914-45

Perfect score: 3463

Sequence: 1 MGSGRDPASVTKNFSL.....RVNMYSSFTMARKGAKDN 648

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-O=/abes/ABSSWEB spool/US10620914/runat\_13032006\_102042\_6770/app query.fasta\_1  
-DB=EST -OPMT=fastap -SUFFIX=p2n.rst -MINMATCH=0\_1 -LOOPECL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abes05h  
-USER=US10620914 @CGN 1 1 5315 @runat\_13032006\_102042\_6770 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:  
1: gb\_est1:  
2: gb\_est2:  
3: gb\_est3:  
4: gb\_est4:  
5: gb\_est5:  
6: gb\_est6:  
7: gb\_est7:  
8: gb\_est8:  
9: gb\_est9:  
10: gb\_est10:  
11: gb\_est11:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1296	37.4	743	5	BU651274
2	1003	29.0	556	1	AV627635
3	847	24.5	475	1	AV644266
4	817	23.6	467	1	AV629528
5	812	23.4	532	1	AV628087
6	696	20.1	391	1	AV626756
7	687	19.8	454	3	BP093208

8	664	19.2	379	3	BP086130
9	643	18.6	508	1	AV641303
10	631	18.2	352	3	BP093875
11	630	18.2	511	1	AV626139
12	627	18.1	354	1	AV631627
13	588	17.0	478	1	AV642589
14	584	16.9	480	1	AV643824
15	556	16.1	466	1	AV628989
16	531	15.3	461	5	BQ824076
17	518	15.0	447	1	AV642307
18	466	13.5	262	1	AV630897
19	418	12.1	398	3	BP093467
20	417	12.0	250	3	BP089550
21	412	11.9	388	3	BP095133
22	401	11.6	377	3	BP097104
23	374	10.8	362	3	BP096956
24	368	10.6	366	3	BP098395
25	353.5	10.2	433	11	AW561229
26	350.5	10.1	838	11	CNS076FM
27	348	10.0	973	11	CNS07804
28	331	9.6	846	8	DR662733
29	325	9.4	287	1	AV640882
30	311	9.0	850	11	CNS06VGM
31	309	8.9	963	11	CNS071FB
32	298	8.6	416	6	CF556716
33	291	8.4	176	1	AV628222
34	252	7.3	800	11	CNS072J0
35	244	7.0	585	3	BI721528
36	233	6.7	320	2	BI189160
37	232.5	6.7	564	1	AJ273357
38	220.5	6.4	1060	10	CW929808
39	219	6.3	301	2	BF072578
40	217.5	6.3	915	11	CNS079G1
41	212.5	6.1	511	9	BZ782470
42	205.5	5.9	817	7	CO023856
43	205.5	5.9	889	7	CO024454
44	202	5.8	877	11	CNS06VXF
45	183	5.3	567	7	CO151218

ALIGNMENTS

RESULT 1

BU651274

LOCUS

DEFINITION

ACCESION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

BU651274 743 bp mRNA linear EST 30-SEP-2002  
1112092D11.y1 C. reinhardtii CC-1690 (mt+), CC-1691 (mt-), Gamete  
(normalized), Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA  
sequence.  
BU651274 GI:23363454  
EST.  
Chlamydomonas reinhardtii  
Chlamydomonas reinhardtii  
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;  
Chlamydomonadales; Chlamydomonadaceae; Chlamydomonads.  
1 (bases 1 to 743)  
Grossman, A., Chang, C.-W., Davies, J., Harris, B., Hauser, C.,  
Lefebvre, P., McDermott, J., Shrager, J., Silflow, C. and Stern, D.  
Analyses of the Chlamydomonas reinhardtii Genome: A Model,  
Unicellular System for Analyzing Gene Function and Regulation in  
Vascular Plants. Project: 1112  
Unpublished (2002)  
Contact: Charles Hauser  
DCMB Box 91000  
Durham, NC 27708-1000  
Tel: 919 613 8159  
Fax: 919 613 8177  
Email: chauser@duke.edu.  
Location/Qualifiers  
1..743  
/organism="Chlamydomonas reinhardtii"  
/mol\_type="mRNA"

/strain="21gr (CC-1690 wild type mt+) & 6145c (CC-1691 wild type mt-)"  
 /db\_xref="taxon:3055"  
 /clone\_lib="C. reinhardtii CC-1690 (mt+), CC-1691 (mt-), Gamete (normalized), Lambda Zap II"  
 /note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2: XhoI; Gamete library was constructed by Hui Zhao, Min Lu, Jeffrey McDermott, William J. Snell and John Davies. Strain 21gr cells (CC-1690; mating type plus) and strain 6145c cells (CC-1691; mating type minus) that had been growing on a light-dark cycle (13:11 L/D) in R-medium (Sager and Granick) were separately transferred into nitrogen-free medium at 8 hours into the light period. PolyA mRNA was purified from each sample every 2 hours for the next 18 hours. The mRNA was pooled and used for cDNA synthesis. The cDNA was directionally cloned into lambda zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the lambda ZAP clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al., (1996) Genome Research 6: 791-806."

## ORIGIN

Alignment Scores:  
 Pred. No.: 6 23e-134 Length: 743  
 Score: 1296.00 Matches: 243  
 Percent Similarity: 99.2% Conservative: 0  
 Best Local Similarity: 99.2% Mismatches: 2  
 Query Match: 37.4% Indels: 1  
 DB: 5 Gaps: 0

US-10-620-914-45 (1-648) x BU651274 (1-743)

Qy 228 GluGlnAsnThrGlnGlySerIleProTyrValProTyrLeuArgAlaProTyrTyrVal 247  
 Db 8 GAGCAGAACACCCAGGCTTCGATCCCTACGTGCGGTGGTGGCGGCCCTACTACGTG 67  
 Qy 248 TrpIleGlyArgLeuProSerValGlyHisAlaLeuHisGluGluArgValGluArgPro 267  
 Db 68 TGGATTGGCGCGCTGCCAGCGTTGGCCAGCCCTGCACGAGGAGCGCGTGGAGCGCGC 127  
 Qy 268 ProMetPheProThrPheLeuTyrThrGlnSerTyrGluAspProGluProAspMet 287  
 Db 128 CCATGTTCCCGCCACCTTCCTGTATACGAGTCTGGGAGGACCCCGAGCGGATATG 187  
 Qy 288 GluValMetGluLeuAsnProLysAspThrValLeuThrLeuThrSerGlyGlyCysAsn 307  
 Db 188 GAGGTGATGGAGATCAACCCCAAGGACACGCGTGTGACCTGACCTAGCGCGGCTGCAAT 247  
 Qy 308 AlaLeuAsnLeuLeuValGlnGlyAlaGlyGlnValValSerValAspCysAsnProAla 327  
 Db 248 GCCTGAACTGCTGGTGGCGAGGGCGCGCGAGGTGTGCGTGGAGTGCACCCCGCG 307  
 Qy 328 GlnSerAlaLeuLeuGluLeuLysValAlaIleGlnGlnLeuGluPheGluAspVal 347  
 Db 308 CAGTCGGCGCTTCGGAGCTGAAGAAGTGGCCATTTCAGCAGCTGGAGTTTGAGCAGTG 367  
 Qy 348 TrpGlnLeuPheGlyGluGlyValHisProArgIleGluGluLeuTyrGluLysLeu 367  
 Db 368 TGGCAGCTGTTCGGCGAGGGCGTGCACCGCGCATTTGAGGAGCTGTACGAGAGAGCTG 427  
 Qy 368 AlaProPheLeuSerGlnThrSerHisAsnPheTrpSerLysArgLeuTyrTyrPheGln 387  
 Db 428 CGCGCCCTTCCTGTCGCAAAACAGACCACTTCCTGGTCCCAAGCGCCCTGTGACTTCAG 487  
 Qy 388 HisGlyLeuTyrTyrGlnGlyGlyMetGlyLysLeuCysTrpValLeuGlnCysLeuAla 407  
 Db 488 CACGCGCTGTACTACAGCGCGGATGGCAAGCTGTGCTGGGTGCTGCAGTGGCTGGCC 547  
 Qy 408 ValValLeuGlyLeuGlyLysThrValLysArgLeuAlaAsnAlaProThrMetGluGlu 427  
 Db 548 GTGGTGTGGGACTGGGCAAGACCGTCAAGCGCCTCGCAACCGCGCCCAATGGAGGAG 607

Qy 428 GlnArgArgLeuTyrAspSerAsnMetLeuIleHisPheValLysAsnGlyProLysPro 447  
 Db 608 CAGCGCGCTGTGGAGCAGCAACATGCTCATTCATCTTGTGAGAACGGGCCCAAGCG 667  
 Qy 448 LeuValTrpLeuPheValLysPheValSerLeuValLeuPheAsnLysAlaValLeuTyr 467  
 Db 668 CTGGTGTGGCTGTTCGTCAAGTTTCGTGAGCTGCTCTTCAACAG-GCGGTGCTGTGG 726  
 Qy 468 PheGlyGlyGlyVal 472  
 Db 727 TTCNCGCGCGCGCTG 741  
 RESULT 2  
 AV627635 556 bp mRNA linear EST 15-DEC-2000  
 LOCUS AV627635 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas  
 DEFINITION reinhardtii cDNA clone LCL029g11\_r 5', mRNA sequence.  
 AV627635  
 ACCESSION AV627635.1 GI:10790269  
 VERSION  
 KEYWORDS EST.  
 SOURCE Chlamydomonas reinhardtii  
 ORGANISM Chlamydomonas reinhardtii  
 Chlamydomonas reinhardtii  
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;  
 Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.  
 REFERENCE 1 (bases 1 to 556)  
 AUTHORS Asamizu, E., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohyama, K.,  
 Nakamura, Y. and Tabata, S.  
 TITLE Generation of expressed sequence tags from low-CO2 and high-CO2  
 adapted cells of Chlamydomonas reinhardtii  
 JOURNAL DNA Res. 7 (5), 305-307 (2000)  
 PUBMED 11089912  
 COMMENT Contact: Erika Asamizu  
 The First Laboratory for Plant Gene Research  
 Kazusa DNA Research Institute  
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.  
 FEATURES  
 Location/Qualifiers  
 1..556  
 /organism="Chlamydomonas reinhardtii"  
 /mol\_type="mRNA"  
 /strain="C9"  
 /db\_xref="taxon:3055"  
 /clone\_lib="Chlamydomonas reinhardtii 5% to 0.04% CO2"  
 /note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2:  
 XhoI; The cDNA library was constructed from cells cultured  
 in a carbon stress acclimatized condition in which carbon  
 dioxide concentration in the bubbling gas was changed from  
 5% to 0.04%"  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 2 67e-101 Length: 556  
 Score: 1003.00 Matches: 184  
 Percent Similarity: 100.0% Conservative: 0  
 Best Local Similarity: 100.0% Mismatches: 0  
 Query Match: 29.0% Indels: 0  
 DB: 1 Gaps: 0  
 US-10-620-914-45 (1-648) x AV627635 (1-556)  
 Qy 96 LysPheLysSerIleTyrValValAspLeuCysHisSerLeuCysGluValAlaLysLys 115  
 Db 3 AAGTTCAAGTCCATCTACGTGGTGGACCTGTGTCACCTCGCTGGAGGTGGCCAAAGAG 62  
 Qy 116 LysAlaLysAlaLysGlyTrpLysAsnValGlnValGluAlaAspAlaCysGlnPhe 135  
 Db 63 AAGCGAAGCCCAAGGCGCTGGAAGAATGTCCAGTCTGTGAGGCGCGCTTGCCTTTT 122  
 Qy 136 AlaProGluGlyThrAlaThrLeuIleThrPheSerTyrSerLeuThrMetIlePro 155  
 Db 123 GCGCCCTTGGAGGACCGCGAGCTCATCACCTTCTCTACTCGCTCAGATGATTCCA 182





## ORIGIN

Alignment Scores: 1.41e-80 Length: 467  
 Pred. No.: 817.00 Matches: 154  
 Score: 89.4% Conservative: 0  
 Percent Similarity: 99.4% Mismatches: 1  
 Best Local Similarity: 23.6% Indels: 0  
 Query Match: 23.6% Gaps: 0  
 DB: 1

US-10-620-914-45 (1-648) x AV629528 (1-467)

QY 60 AlaAlaPheAlaAlaArgLeuAlaGluArgSerAsnLeuIleTrrValAspLeuGlyGly 79  
 DB 1 GCTGCAGTTCCTGCCCGCTGGCGAGCGCTCGAACCTCATCTGGGTTCACCTGGGTGGT 60  
 QY 80 GlyThrGlyGluAsnValAspMetMetAlaAspTrrIleAspLeuAlaLysPheLysSer 99  
 DB 61 GGCACTGGGAGAGATTCGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120  
 QY 100 IleTrrValValAspLeuCysHisSerLeuCysGluValAlaLysLysLysLysAlaLysAla 119  
 DB 121 ATCTACGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180  
 QY 120 LysGlyTrrPlysAsnValGlnValValGluAlaAspAlaCysGlnPheAlaProProGlu 139  
 DB 181 AAGGCTGGGAAGATTCAGAGTTCGAGGCGCGAGCGCTTGCCTGCTGCTGCTGCTGCTGCTGCTGCT 240  
 QY 140 GlyThrAlaThrIleuLeuThrPheSerTrrSerLeuThrMetIleProProPheHisAsn 159  
 DB 241 GGCACCGCGAGCTCATCATCT 300  
 QY 160 ValIleAspGlnAlaCysSerTrrLeuSerGlnAspGlyLeuValGlyValAlaAlaAspPhe 179  
 DB 301 GTATGACGACGAGCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360  
 QY 180 TrrValSerGlyLysTrrAspLeuProLeuArgGlnMetProTrrSerArgArgPhePhe 199  
 DB 361 TACGTGAGCGGCAAGTACGACCTGCCCTGCGCAGATGCCCTGGTGGCGGCTTCTCTCTCTCTCTCTCT 420  
 QY 200 TrrArgSerIlePheAspIleAspAsnIleAspIleGlyProGlu 214  
 DB 421 TGGCGATCGATCTTCGACATCGACCAACATTCGACATCGGCGCGCGAG 465

RESULT 5  
 AV628087 532 bp mRNA linear EST 15-DEC-2000  
 LOCUS AV628087 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas  
 DEFINITION reinhardtii cDNA clone LCL036e03\_r 5', mRNA sequence.

AV628087 GI:10790721

AV628087

EST.

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBLISHED

COMMENT

CONTACT: Erika Asamizu

The First Laboratory for Plant Gene Research

Kazusa DNA Research Institute

Yana 1532-3, Kisarazu, Chiba 292-0812, Japan

/db\_xref="taxon:3055"

/clone\_lib="Chlamydomonas reinhardtii 5% to 0.04% CO2"

/note="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:

XhoI; The cDNA library was constructed from cells cultured

in a carbon stress acclimatized condition in which carbon

dioxide concentration in the bubbling gas was changed from

5% to 0.04%"

## ORIGIN

Alignment Scores: 6.32e-80 Length: 532  
 Pred. No.: 812.00 Matches: 158  
 Score: 89.8% Conservative: 0  
 Percent Similarity: 89.3% Mismatches: 1  
 Best Local Similarity: 23.4% Indels: 18  
 Query Match: 23.4% Gaps: 1  
 DB: 1

US-10-620-914-45 (1-648) x AV628087 (1-532)

QY 3 SerGlyArgAspGlyArgProAlaSerTrrThrLysLysAsnPheSerLeuGluLysLeu 22  
 DB 1 TCGGTCGTGACGCGCGCTGCGAGCTACACCAAGAGACTTCTCCCTGGAGAGCTC 60  
 QY 23 LysLeuSerSerMetLysAspLeuThrValLeuArgHisMetTrrPheGlySerLys 42  
 DB 61 AAGCTCAGCAGCATGAAGGATGACCTGACCGTTCTGCGCCATATGTGTTTCGCGCAGCAAG 120  
 QY 43 LysGlyAspAspHisAlaAlaArgLeuGluSerPheTrrGlyProGlnAlaAlaAlaPhe 62  
 DB 121 AAGGCGGATGATCACGCTGCTGCTGCGAGAGCTTCTACGGGCGCCCGCGCTGCTGCTAT 180  
 QY 63 -----AlaAla 64  
 DB 181 GATGCTTTCGCTGCGGTTCTCTCGGGTCGCGAGCCCATGCTCGTGCAGTTGCTGCC 240  
 QY 65 ArgLeuAlaGluArgSerAsnLeuIleTrrValAspLeuGlyGlyThrGlyGluAsn 84  
 DB 241 CGCCTGGCGAGCGCTCGAACCTCATCTGGGTTGACCTGGGTTGGTGGCACTGGGGAGAA 300  
 QY 85 ValAspMetMetAlaAspTrrIleAspLeuAlaLysPheLysSerIleTrrValValAsp 104  
 DB 301 GTCGATATGATGGCTGATTACATGACCTGGCGAGTTCAAGTCCATCTACGTGGTGCAC 360  
 QY 105 LeuCysHisSerLeuCysGluValAlaLysLysLysLysLysLysLysLysLysLysLys 124  
 DB 361 CTGTGCACCTCGCTGTCGAGGTGGCCAGAGAGAGCGCAAGGCGCAAGGCTGGAGAA 420  
 QY 125 ValGlnValValGluAlaAspAlaCysGlnPheAlaProProGluGlyThrAlaThrLeu 144  
 DB 421 GTCCAGTCTGTGGAGCGGCGCTTGCCTGCAATTTGCGCCCTCGAGGCGACCGGACGCTC 480  
 QY 145 IleThrPheSerTrrSerLeuThrMetIleProProPheHisValIle 161  
 DB 481 ATACCTTCTCTCTACTCGTTCACGATGATTCACCGCTTCCACCAAGCTCATC 531

RESULT 6

AV626756

LOCUS

DEFINITION

AV626756

AV626756

EST.

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

AV626756 391 bp mRNA linear EST 15-DEC-2000  
 AV626756 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas  
 reinhardtii cDNA clone LCL014h09\_r 5', mRNA sequence.

AV626756

AV626756.1 GI:10789036

EST.

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

Generation of expressed sequence tags from low-CO2 and high-CO2  
 adapted cells of Chlamydomonas reinhardtii

Contact: Erika Asamizu

The First Laboratory for Plant Gene Research

Kazusa DNA Research Institute

Yana 1532-3, Kisarazu, Chiba 292-0812, Japan

Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

Location/Qualifiers

1. 532

/organism="Chlamydomonas reinhardtii"

/mol\_type="mRNA"

/strain="C9"

JOURNAL DNA Res. 7 (5), 305-307 (2000)  
 PUBMED 11089912  
 COMMENT Contact: Erika Asamizu  
 The First Laboratory for Plant Gene Research  
 Kazusa DNA Research Institute  
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES  
 source  
 1. .391  
 Location/Qualifiers  
 /organism="Chlamydomonas reinhardtii"  
 /mol\_type="mRNA"  
 /strain="C9"  
 /db\_xref="taxon:3055"  
 /clone\_lib="ICL014h09\_r"  
 /clone\_lib="Chlamydomonas reinhardtii 5% to 0.04% CO2"  
 /note="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2: XhoI; The cDNA library was constructed from cells cultured in a carbon stress acclimatized condition in which carbon dioxide concentration in the bubbling gas was changed from 5% to 0.04%"

ORIGIN  
 Alignment Scores:  
 Pred. No.: 3 85e-67 Length: 391  
 Score: 696.00 Matches: 130  
 Percent Similarity: 100.0% Conservative: 0  
 Best Local Similarity: 100.0% Mismatches: 0  
 Query Match: 20.1% Indels: 0  
 DB: 1 Gaps: 0

US-10-620-914-45 (1-648) x AV626756 (1-391)

QY 77 LeuGlyGlyGlyThrGlyGluAenValAspMetAlaAspTyrIleAspLeuAlaLys 96  
 DB 1 CTGGGTGGTGGCACTGGCGAGATGTCGATATGCTGATTACATCGACCTGGCGAAG 60

QY 97 PheLysSerIleTyrValValAspLeuCysHisSerLeuCysGluValAlaLysLys 116  
 DB 61 TTCAGTCCATCTACGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 120

QY 117 AlaLysAlaLysGlyTyrLysAenValGlnValValAlaAspAlaCysGlnPheAla 136  
 DB 121 GCGAGGCCAAGGCGTGGAGAGATGCCAGTCTGGAGCGCCGACCTTGCCTAATTGGC 180

QY 137 ProProGluGlyThrAlaThrLeuIleThrPheSerTyrSerLeuThrMetIleProPro 156  
 DB 181 CCCCTGAGGGCACCGGAGCTCATCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 240

QY 157 PheHisAenValIleAspGlnAlaCysSerTyrLeuSerGlnAspGlyLeuValGlyVal 176  
 DB 241 TTCACACAGCTCATCGACCGAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 300

QY 177 AlaAspPheTyrValSerGlyLysTyrAspLeuProLeuArgGlnMetProTyrSerArg 196  
 DB 301 GCGAGCTTCTACGTAGCGGCAAGTACGACCTGCTGCTGCTGCTGCTGCTGCTGCTG 360

QY 197 ArgPhePheTyrArgSerIlePheAspIle 206  
 DB 361 CGTTTCTTCTGGCGATCGATCTTCGACATC 390

RESULT 7  
 BP093208  
 LOCUS BP093208 Chlamydomonas reinhardtii cDNA clone MXL005a04\_r 5', mRNA sequence.  
 DEFINITION Chlamydomonas reinhardtii cDNA clone MXL005a04\_r 5', mRNA sequence.  
 ACCESSION BP093208  
 VERSION BP093208.1 GI:49465295  
 KEYWORDS EST.  
 SOURCE Chlamydomonas reinhardtii  
 ORGANISM Chlamydomonas reinhardtii  
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;  
 Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.  
 1 (bases 1 to 454)

AUTHORS Asamizu, E., Nakamura, Y., Miura, K., Fukuzawa, H., Fujiwara, S.,  
 Hirono, M., Iwamoto, K., Matsuda, Y., Minagawa, J., Shimogawara, K.,  
 Takahashi, Y. and Tabata, S.  
 TITLE Establishment of Publicly Available cDNA Material and Information  
 Resource of Chlamydomonas reinhardtii (Chlorophyta), to Facilitate  
 Gene Function Analysis  
 JOURNAL Phycologia (2004) in press  
 COMMENT The First Laboratory for Plant Gene Research  
 Kazusa DNA Research Institute  
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES  
 source  
 1. .454  
 Location/Qualifiers  
 /organism="Chlamydomonas reinhardtii"  
 /mol\_type="mRNA"  
 /strain="C9"  
 /db\_xref="taxon:3055"  
 /clone\_lib="MXL005a04\_r"  
 /clone\_lib="Chlamydomonas reinhardtii C9 various conditions"  
 /note="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2: XhoI; The cDNA library was made from a mixture of cells grown under various conditions"

ORIGIN  
 Alignment Scores:  
 Pred. No.: 4 99e-66 Length: 454  
 Score: 687.00 Matches: 132  
 Percent Similarity: 88.1% Conservative: 1  
 Best Local Similarity: 87.4% Mismatches: 0  
 Query Match: 19.8% Indels: 18  
 DB: 3 Gaps: 1

US-10-620-914-45 (1-648) x BP093208 (1-454)

QY 36 HisMetTyrPheGlySerLysGlyAspAspHisAlaAlaArgLeuGluSerPheTyr 55  
 DB 1 CATATGTGTGGTGGCAGCAAGAGGCGATGATCATCGTCTGGCTGGAGAGCTTCTAC 60

QY 56 GlyProGlnAlaAlaAlaPhe----- 62  
 DB 61 GGGCCCCAGGCGCTGCTTATGATGCTTCCGGTGGGCTTCTCTGGGGTGGCAGGCC 120

QY 63 -----AlaAlaArgLeuAlaGluArgSerAsnLeuIleTyrValAspLeu 77  
 DB 121 ATGCTCGTGCAGTGTGCTGCCCGCTGGCGAGCGCTCGAACCCTCATCTGGGTGACCTG 180

QY 78 GlyGlyGlyThrGlyGluAenValAspMetMetAlaAspTyrIleAspLeuAlaLysPhe 97  
 DB 181 GGTGGTGGCACTGGGAGAAATGTCGATATGATGGCTGATTACATCGACCTGGCGAAGTTC 240

QY 98 LysSerIleTyrValValAspLeuCysHisSerLeuCysGluValAlaLysLysAla 117  
 DB 241 AAGTCCATCTACGTGTGCTGACCTGCGCCACTCGCTGTGGAGGTGGCCGAAGAGAGCG 300

QY 118 LysAlaLysGlyTyrLysAenValGlnValValAlaAspAlaCysGlnPheAlaPro 137  
 DB 301 AAGGCCAAGGCTGGAGAAATGTCAGTCTGGAGGCCGACGCTGCCAATTTGGGCC 360

QY 138 ProGluGlyThrAlaThrLeuIleThrPheSerTyrSerLeuThrMetIleProPhe 157  
 DB 361 CCGTGGGCGACCGCGACGCTCATCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 420

QY 158 HisAenValIleAspGlnAlaCysSerTyrLeu 168  
 DB 421 CACAAAGTTCATCGACCGAGCTTGTCTGTACCTG 453

RESULT 8  
 BP086130  
 LOCUS BP086130 Chlamydomonas reinhardtii C9 various conditions  
 DEFINITION Chlamydomonas reinhardtii cDNA clone MX002e03\_r 5', mRNA sequence.

```

ACCESSION BP086130
VERSION BP086130.1 GI:49458217
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii
ORGANISM Chlamydomonas reinhardtii
REFERENCE 1 (bases 1 to 379)
AUTHORS Asamizu,E., Nakamura,Y., Miura,K., Fukuzawa,H., Fujiwara,S.,
Hirono,M., Iwanoto,K., Matsuda,Y., Minagawa,J., Shimogawara,K.,
Takahashi,Y. and Tabata,S.
TITLE Establishment of Publicly Available cDNA Material and Information
Resource of Chlamydomonas reinhardtii (Chlorophyta), to Facilitate
Gene Function Analysis
JOURNAL Phycologia (2004) In press
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
source
1..379
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="C9"
/db_xref="taxon:3055"
/clone="WX002e03_r"
/notes="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was made from a mixture of cells
grown under various conditions"

ORIGIN
Alignment Scores:
Pred. No.: 1,41e-63 Length: 379
Score: 664.00 Matches: 126
Percent Similarity: 100.0% Conservatives: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 19.2% Indels: 0
DB: 3 Gaps: 0

US-10-620-914-45 (1-648) x BP086130 (1-379)
QY 484 GlyIleProGluAsnTyrIleAlaArgThrMetAspGlyValAlaGluAsnSerHis 503
DB 2 GGCAATCCCAATTGAGAACTACATCGCGCGCACCATGGACCGCGTGGCGGAGAACTCGCAC 61
QY 504 ValArgLysGlnAsnTyrPheTyrTyrAsnCysLeuThrGlyLysPheLeuArgAspAsn 523
DB 62 GTGGCGAAGCAGAACTACTTCTACTACAACTGCTCACCGGCAAGTTCCTTGGCGGACAC 121
QY 524 CysProThrTyrLeuArgGluAlaAlaPheAlaThrLeuLysSerGlyValValAspAsn 543
DB 122 TGCCCCACCTTACCTGCGGAGGCGGCTTCCGCCACCCTCAAGAGTGGCGTGGTGGACAC 181
QY 544 LeuThrValSerThrAsnPheMetGluLeuLysAlaArgThrTyrThrLysVal 563
DB 182 CTGACCGCTCTCCAACTTTCTTCATGGAGGAGCTCAAGGCGGCACCTACACCAAGGTG 241
QY 564 IleLeuMetAspHisValAspTyrLeuAspMetProValAlaAsnGluLeuAlaGluCys 583
DB 242 ATTCTGATGACCAACGCTGGACTGGCTGGATATGCGCTGGCCCAACGAGCTGGCGGAGTGC 301
QY 584 LeuAlaLysGlnValAlaProGlyGlyValIleValIleTyrArgSerAlaSerLeuSerPro 603
DB 302 CTGGCCCAAGCAGGTGGCGCGGCGGCATCGTCATCTGGCGCTCCGCTCCCTCAGCCCG 361
QY 604 ProTyrAlaGluLeuIle 609
DB 362 CCCTACCGCCGAGCTGATC 379

RESULT 9

```

---

```

AV641303
LOCUS AV641303 Chlamydomonas reinhardtii 588 bp mRNA linear EST 15-DEC-2000
DEFINITION AV641303 Chlamydomonas reinhardtii 588 bp mRNA sequence.
ACCESSION AV641303
VERSION AV641303.1 GI:10784631
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii
ORGANISM Chlamydomonas reinhardtii
REFERENCE 1 (bases 1 to 508)
AUTHORS Asamizu,E., Miura,K., Kucho,K., Inoue,Y., Fukuzawa,H., Ohyama,K.,
Nakamura,Y. and Tabata,S.
TITLE Generation of expressed sequence tags from low-CO2 and high-CO2
adapted cells of Chlamydomonas reinhardtii
JOURNAL DNA Res. 7 (5), 305-307 (2000)
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
source
1..508
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="C9"
/db_xref="taxon:3055"
/clone="HCL031d02_r"
/notes="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was constructed from cells cultured
in a medium with bubbling air containing 5% carbon
dioxide"

ORIGIN
Alignment Scores:
Pred. No.: 5,1e-61 Length: 508
Score: 643.00 Matches: 126
Percent Similarity: 87.6% Conservatives: 1
Best Local Similarity: 86.9% Mismatches: 0
Query Match: 18.6% Indels: 18
DB: 1 Gaps: 1

US-10-620-914-45 (1-648) x AV641303 (1-508)
QY 1 MetGlySerGlyArgAspGlyArgProAlaSerTyrThrLysLysAsnPheSerLeuGlu 20
DB 72 ATGGGGTGGGTCTGTCAGCGCGCGCTCGAGCTACACCAAGAAGAACTTCTCCCTGGAG 131
QY 21 LysLeuLysLeuSerSerMetLysAspAspLeuThrValLeuArgHisMetTyrPheGly 40
DB 132 AAGCTCAAGCTCAGCAGCATGAAGGATGACCTGACCGTCTCGGCCCATATGTGGTTCGGC 191
QY 41 SerLysLysGlyAspAspHisAlaAlaArgLeuGluSerPheTyrGlyProGlnAlaAla 60
DB 192 AGCAAGAGGCGCATGATCATCGCTGCTCGCTGGAGAGCTTCTACGGGCCCCAGCCGCT 251
QY 61 AlaPhe----- 62
DB 252 GCCTATGATGCTTTTCCGGTCCGCTTCTCTCGGGTTCGAGGCCCATGCTCGCTGACGTT 311
QY 63 AlalaArgLeuAlaGluArgSerAsnLeuIleTyrValAspLeuGlyGlyThrGly 82
DB 312 GCTGCCCGCTGGCGGCGCTCGAACCTCATCTGGGTTCACCTGGGTGGTGGCAGCTGGG 371
QY 83 GluAsnValAspMetMetAlaAspTyrIleAspLeuAlaLysPheLysSerIleTyrVal 102
DB 372 GAGATGTCGATGATGATGGCTGATCATCGACTCGAGTTCAGTTCATCTACGTG 431
QY 103 ValAspLeuCysHisSerLeuGluValAlaLysLysLysAlaLysGlyTyr 122

```

Db 432 GTCGACCTGTGCGCACTCGCTGTGCGAGTGCGCCAAAGAGGCGCAAGGCGCTGG 491

Qy 123 LysAsnValGlnVal 127  
|||||

Db 492 AAGAATGTCCAGGTC 506  
|||||

RESULT 10  
BP093875 352 bp mRNA linear EST 30-JUN-2004  
LOCUS BP093875 Chlamydomonas reinhardtii C9 various conditions  
DEFINITION Chlamydomonas reinhardtii cDNA clone MXL015a05\_r 5', mRNA sequence.  
ACCESSION BP093875  
VERSION BP093875.1 GI:49465962  
KEYWORDS EST.  
SOURCE Chlamydomonas reinhardtii  
ORGANISM Chlamydomonas reinhardtii  
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;  
Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.  
REFERENCE 1 (bases 1 to 352)  
AUTHORS Asamizu, E., Nakamura, Y., Miura, K., Fukuzawa, H., Fujiwara, S.,  
Hirono, M., Iwamoto, K., Matsuda, Y., Minagawa, J., Shimogawara, K.,  
Takahashi, Y. and Tabata, S.  
TITLE Establishment of Publicly Available cDNA Material and Information  
Resource of Chlamydomonas reinhardtii (Chlorophyta), to Facilitate  
Gene Function Analysis  
JOURNAL Psychologia (2004) In press  
COMMENT Contact: Erika Asamizu  
The First Laboratory for Plant Gene Research  
Kazusa DNA Research Institute  
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES  
source  
1. 352  
/organism="Chlamydomonas reinhardtii"  
/mol\_type="mRNA"  
/strain="C9"  
/db\_xref="taxon:3055"  
/clone="MXL015a05\_r"  
/note="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:  
XhoI; The cDNA library was made from a mixture of cells  
grown under various conditions"

ORIGIN  
Alignment Scores:  
Pred. No.: 6.26e-60 Length: 352  
Score: 631.00 Matches: 117  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 18.2% Indels: 0  
DB: 3 Gaps: 0

US-10-620-914-45 (1-648) x BP093875 (1-352)

Qy 420 AlaAsnAlaProThrMetGluGlnArgLeuTrpAspSerAsnMetLeuLeuHis 439  
|||||

Db 2 GCCAACGCGCCCAATATGAGGAGGACGCGCGTCTGTGGGACAGCAATGCTCATCCAC 61  
|||||

Qy 440 PheValLysAnglyProLysProLeuValTrpLeuPheValLysPheValSerLeuVal 459  
|||||

Db 62 TTCGTGAAGAACGCGGCCCAAGCCGCTGTGTGGCTGTTCGTCAGTTCTGTGAGCGCTGGTG 121  
|||||

Qy 460 LeuPheAsnLysAlaValLeuTrpPheGlyGlyValProGlyLysGlnTrpAlaLeu 479  
|||||

Db 122 CTCTTCAACAGGCCGCTGCTGTGTTCGGCGCGCGCTGCCGCGCAGACGATGAGCGCTG 181  
|||||

Qy 480 IleLysAlaAspGlyIleProIleGluAsnTrpIleAlaArgThrMetAspGlyValAla 499  
|||||

Db 182 ATCAAGCGGACGCGCATCCCATTTGAGAACTACATCGCGCGCACCATGAGCGCGTGGCG 241  
|||||

Qy 500 GluAsnSerHisValArgLysGlnAsnTrpPheTrpTyrAsnCysLeuThrGlyLysPhe 519  
|||||

Db 242 GAGAACTCGCAGCTGCGCAAGCAGAACTACTTCTACTACAACTGCTCACCAGCAAGTTC 301

Qy 520 LeuArgAspAsnCysProThrTyrLeuArgGluAlaAlaPheAlaThrLeu 536  
|||||

Db 302 CTGGCGGACAACTGCCCACTTACCTGCGGAGGCGGCTTGGCCACCCTC 352  
|||||

RESULT 11  
AV626139 511 bp mRNA linear EST 15-DEC-2000  
LOCUS AV626139 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas  
DEFINITION reinhardtii cDNA clone LCL003a03\_r 5', mRNA sequence.  
ACCESSION AV626139  
VERSION AV626139.1 GI:10788419  
KEYWORDS EST.  
SOURCE Chlamydomonas reinhardtii  
ORGANISM Chlamydomonas reinhardtii  
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;  
Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.  
REFERENCE 1 (bases 1 to 511)  
AUTHORS Asamizu, E., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohyama, K.,  
Nakamura, Y. and Tabata, S.  
TITLE Generation of expressed sequence tags from low-CO2 and high-CO2  
adapted cells of Chlamydomonas reinhardtii  
JOURNAL DNA Res. 7 (5), 305-307 (2000)  
PUBMED 11089912  
COMMENT Contact: Erika Asamizu  
The First Laboratory for Plant Gene Research  
Kazusa DNA Research Institute  
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES  
source  
1. 511  
/organism="Chlamydomonas reinhardtii"  
/mol\_type="mRNA"  
/strain="C9"  
/db\_xref="taxon:3055"  
/clone="LCL003a03\_r"  
/note="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:  
XhoI; The cDNA library was constructed from cells cultured  
in a carbon stress acclimatized condition in which carbon  
dioxide concentration in the bubbling gas was changed from  
5% to 0.04%"

ORIGIN  
Alignment Scores:  
Pred. No.: 1.47e-59 Length: 511  
Score: 630.00 Matches: 123  
Percent Similarity: 87.3% Conservative: 1  
Best Local Similarity: 86.6% Mismatches: 0  
Query Match: 18.2% Indels: 18  
DB: 1 Gaps: 1

US-10-620-914-45 (1-648) x AV626139 (1-511)

Qy 1 MetGlySerGlyArgAspGlyArgProAlaSerTyrThrLysLysAsnSerLeuGlu 20  
|||||

Db 84 ATGGGCTGGGTCTGTCAGCGCGGCTGCGAGCTACACCAAGAAGAACTTCTCCCTGGAG 143  
|||||

Qy 21 LysLeuLysLeuSerSerMetLysAspAspLeuThrValLeuArgHisMetTrpPheGly 40  
|||||

Db 144 AAGCTCAAGCTCAGCAGCATGAAGATGACCTGACCGCTTCTGCGCCATATATGTGTTCGGC 203  
|||||

Qy 41 SerLysLysGlyAspAspHisAlaAlaArgLeuGluSerPheTyrGlyProGlnAlaAla 60  
|||||

Db 204 AGCAAGAGGCGGATGATCAGCTGCTGCGCTGAGAGCTTCTACGGGCCCCAGCGCGCT 263  
|||||

Qy 61 AlaPhe----- 62  
|||:::

Db 264 GCCTATGATGCTTTCGGTTCGGTTCCTTCGGGTCGAGGCCCATGCTCGCTGCAGTT 323  
|||||

Qy 63 AlaAlaArgLeuAlaGluArgSerAsnLeuIleTrpValAspLeuGlyGlyThrGly 82  
|||||

Db 324 GCTGCCCGCTGGCCGAGCGCTCGAACCTCATCTGGGTTGACCTGGGTGGTGGCACTGGG 383  
 Qy 83 GluValAspMetMetAlaAspTyrIleAspLeuAlaLysPheLysSerIleTyrVal 102  
 Db 384 GAGAATGTCATATGATGGCTGATTACATGACCTGGCGAAGTTCAAGTCCATCTACGTG 443  
 Qy 103 ValAspLeuCyHisSerLeuCyGluValAlaLysLysLysAlaLysGlyTyr 122  
 Db 444 GTGACCTGTGCCACTCGCTGTGCGAGTGTGCCAAGAGGCGAAGGCGCTGG 503  
 Qy 123 LysAsn 124  
 Db 504 AAGAAT 509

## RESULT 12

AV631627 354 bp mRNA linear EST 15-DEC-2000  
 LOCUS AV631627 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas  
 DEFINITION reinhardtii cDNA clone LCL097c01\_r 5', mRNA sequence.  
 ACCESSION AV631627  
 VERSION AV631627.1 GI:10794261  
 KEYWORDS EST.  
 SOURCE Chlamydomonas reinhardtii  
 ORGANISM Chlamydomonas reinhardtii  
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;  
 Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.  
 REFERENCE 1 (bases 1 to 354)  
 AUTHORS Asamizu, E., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohyama, K.,  
 Nakamura, Y. and Tabata, S.  
 TITLE Generation of expressed sequence tags from low-CO2 and high-CO2  
 adapted cells of Chlamydomonas reinhardtii  
 JOURNAL DNA Res. 7 (5), 305-307 (2000)  
 PUBMED 11089912  
 COMMENT Contact: Erika Asamizu  
 The First Laboratory for Plant Gene Research  
 Kazusa DNA Research Institute  
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

## FEATURES

source  
 1..354  
 /organism="Chlamydomonas reinhardtii"  
 /mol\_type="mRNA"  
 /strain="C9"  
 /db\_xref="taxon:3055"  
 /clone="LCL097c01\_r"  
 /clone\_lib="Chlamydomonas reinhardtii 5% to 0.04% CO2"  
 /note="vector: pBluescriptII SK-; Site: 1: EcoRI; Site 2:  
 XhoI; The cDNA library was constructed from cells cultured  
 in a carbon stress acclimatized condition in which carbon  
 dioxide concentration in the bubbling gas was changed from  
 5% to 0.04%"

## ORIGIN

Alignment Scores:  
 Pred. No.: 1.77e-59 Length: 354  
 Score: 627.00 Matches: 117  
 Percent Similarity: 100.0% Conservative: 0  
 Best Local Similarity: 100.0% Mismatches: 0  
 Query Match: 18.1% Indels: 0  
 DB: 1 Gaps: 0

US-10-620-914-45 (1-648) x AV631627 (1-354)

Qy 111 GluValAlaLysLysLysAlaLysGlyTyrLysAsnValGlnValGluAla 130  
 Db 3 GAGGTGGCCAGAGAAGGCGAAGCGGCTGGGAAGATGTCCAGGTGCTGGAGGCC 62  
 Qy 131 AspAlaCyGlnPheAlaProGluGlyThrAlaThrLeuIleThrPheSerTyrSer 150  
 Db 63 GACGCTTGCCAAATTTGGCCCCCTTGAGGGCACCGGACGCTCATCCTTCTCTACTCG 122  
 Qy 151 LeuThrMetIleProPheHisAsnValIleAspGlnAlaCySerTyrLeuSerGln 170

Db 123 CTCACGATGATTCCACCGCTCCACACGTCATCGACAGCGTTGCTGCTACCTGTCCCAA 182  
 Qy 171 AspGlyLeuValGlyValAlaAspPheTyrValSerGlyLysTyrAspLeuProLeuArg 190  
 Db 183 GACGGCTCTGGTGGCGGTTTCCGCACTTCTACGTGAGCGGCAAGTACGACCTGCCCTGGCG 242  
 Qy 191 GlnMetProTyrSerArgPhePheTyrArgSerIlePheAspIleAspAsnIleAsp 210  
 Db 243 CAGATGCCCTGGTGGCGCGCTTCTTCTGGCGATCGATCTTCGACATCGACAACATTGAC 302  
 Qy 211 IleGlyProGluArgAlaTyrLeuGluGlnLysLeuGluArgValTyr 227  
 Db 303 ATCGGGCCCGAGCGCGCGCTACCTGGAGCAGAAGCTGGAGCGCGTGTGG 353

## RESULT 13

AV642589 478 bp mRNA linear EST 15-DEC-2000  
 LOCUS AV642589 Chlamydomonas reinhardtii 5% CO2 Chlamydomonas reinhardtii  
 DEFINITION cDNA clone HCL054c11\_r 5', mRNA sequence.  
 ACCESSION AV642589  
 VERSION AV642589.1 GI:10785917  
 KEYWORDS EST.  
 SOURCE Chlamydomonas reinhardtii  
 ORGANISM Chlamydomonas reinhardtii  
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;  
 Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.  
 REFERENCE 1 (bases 1 to 478)  
 AUTHORS Asamizu, E., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohyama, K.,  
 Nakamura, Y. and Tabata, S.  
 TITLE Generation of expressed sequence tags from low-CO2 and high-CO2  
 adapted cells of Chlamydomonas reinhardtii  
 JOURNAL DNA Res. 7 (5), 305-307 (2000)  
 PUBMED 11089912  
 COMMENT Contact: Erika Asamizu  
 The First Laboratory for Plant Gene Research  
 Kazusa DNA Research Institute  
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

## FEATURES

source  
 1..478  
 /organism="Chlamydomonas reinhardtii"  
 /mol\_type="mRNA"  
 /strain="C9"  
 /db\_xref="taxon:3055"  
 /clone="HCL054c11\_r"  
 /clone\_lib="Chlamydomonas reinhardtii 5% CO2"  
 /note="vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:  
 XhoI; The cDNA library was constructed from cells cultured  
 in a medium with bubbling air containing 5% carbon  
 dioxide"

## ORIGIN

Alignment Scores:  
 Pred. No.: 6.74e-55 Length: 478  
 Score: 588.00 Matches: 116  
 Percent Similarity: 86.7% Conservative: 1  
 Best Local Similarity: 85.9% Mismatches: 0  
 Query Match: 17.0% Indels: 18  
 DB: 1 Gaps: 1

US-10-620-914-45 (1-648) x AV642589 (1-478)

Qy 1 MetGlySerGlyArgAspGlyArgProAlaSerTyrThrLysLysAsnPheSerLeuGlu 20  
 Db 74 ATGGGGTGGGGTGGTGGCGCCCGCTCGAGGTACACCAAGAGAACTTCTCCCTGGAG 133  
 Qy 21 LysLeuLysLeuSerSerMetLysAspAspLeuThrValLeuArgHisMetTrpPheGly 40  
 Db 134 AAGCTCAAGCTCAGCAGCATGAAGGATGACCTGACCGTCTCGGCCATATATGTGTTGGC 193  
 Qy 41 SerLysLysGlyAspAspHisAlaAlaArgLeuGluSerPheTyrGlyProGlnAlaAla 60  
 Db 194 AGCAAGAAGGCGCATGATCATCGCTGCTGCTGAGAGCTTCTACGGGCCCCCGCCGCT 253



```

Qy 61 AlaPhe-----62
Db 254 GCCTATGATGCTTTCCGGTCGCGGTTCTCTGGGGTCGAGGCCCATGCTCGCTGCAGTT 313
Qy 63 AlaAaArgLeuAlaGluArgSerAsnLeuIleTrpValAspLeuGlyGlyThrGly 82
Db 314 GCTGCCCGCTGGCCGAGCGCTCGAACCTCATCTGGGTTGACCTGGGTGGTGGCACTGGG 373
Qy 83 GluAsnValAspMetMetAlaAspTyrIleAspLeuAlaLysPheLysSerIleTyrVal 102
Db 374 GAGATGTGATATGATGGCTGATTACATCGACTGGCGAAGTTCAAGTCCATCTACGTG 433
Qy 103 ValAspLeuGlyHisSerLeuGluValAlaLysLysAla 117
Db 434 GTCGACCTGTGCCACTCGCTGTCGAGTGCAGGTCGCAAGAAGCGC 478

RESULT 14
LOCUS AV643824 480 bp mRNA linear EST 15-DEC-2000
DEFINITION AV643824 Chlamydomonas reinhardtii 5% CO2 Chlamydomonas reinhardtii
cDNA clone HCL076g02_r 5', mRNA sequence.
ACCESSION AV643824.1 GI:10787152
VERSION AV643824.1
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;
Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.
REFERENCE 1 (bases 1 to 480)
AUTHORS Asamizu, E., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohyama, K.,
Nakamura, Y. and Tabata, S.
TITLE Generation of expressed sequence tags from low-CO2 and high-CO2
adapted cells of Chlamydomonas reinhardtii
JOURNAL DNA Res. 7 (5), 305-307 (2000)
PUBMED 11089912
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
source
1..480
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="C9"
/db_xref="taxon:3055"
/clone="HCL076g02_r"
/notes="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was constructed from cells cultured
in a medium with bubbling air containing 5% carbon
dioxide"

ORIGIN
Alignment Scores: 1.91e-54 Length: 480
Pred. No.: 584.00 Matches: 115
Score: 86.6% Conservative: 1
Percent Similarity: 85.8% Mismatches: 0
Best Local Similarity: 16.9% Indels: 18
Query Match: 1
Gaps: 1
DB: 1

US-10-620-914-45 (1-648) x AV643824 (1-480)

Qy 1 MetGlySerGlyArgAspGlyArgProAlaSerTyrThrLysLysAsnPheSerLeuGlu 20
Db 77 ATGGGGTGGGTGCTGTCAGCGCGCGCTCGAGCTACACCAAGAAGAACTTCTCCCTGGAG 136
Qy 21 LysLeuLysLeuSerSerMetLysAspAspLeuThrValLeuArgHisMetTrpPheGly 40
Db 137 AGCTCAAGCTCAGCAGCATGAAGGATGACCTGACCGCTTCTGCGCCATATGTGTGGC 196

```

```

Qy 41 SerLysGlyAspAspHisAlaAlaArgLeuGluSerPheTyrGlyProGlnAlaAla 60
Db 197 AGCAAGAAGCGCGATGATCATCGCTGCTGCTGAGAGCTTCTACGGGCCCGCGCT 256
Qy 61 AlaPhe-----62
Db 257 GCCTATGATGCTTTCCGGTCGCGGTTCTCTGGGGTCGAGGCCCATGCTCGCTGCAGTT 316
Qy 63 AlaAaArgLeuAlaGluArgSerAsnLeuIleTrpValAspLeuGlyGlyThrGly 82
Db 317 GCTGCCCGCTGGCCGAGCGCTCGAACCTCATCTGGGTTGACCTGGGTGGTGGCACTGGG 376
Qy 83 GluAsnValAspMetMetAlaAspTyrIleAspLeuAlaLysPheLysSerIleTyrVal 102
Db 377 GAGATGTGATATGATGGCTGATTACATCGACTGGCGAAGTTCAAGTCCATCTACGTG 436
Qy 103 ValAspLeuGlyHisSerLeuGluValAlaLysLysAla 116
Db 437 GTCGACCTGTGCCACTCGCTGTCGAGTGCAGGTCGCAAGAAG 478

RESULT 15
LOCUS AV628989 466 bp mRNA linear EST 15-DEC-2000
DEFINITION AV628989 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas
reinhardtii cDNA clone LCL050a06_r 5', mRNA sequence.
ACCESSION AV628989
VERSION AV628989.1 GI:10791623
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;
Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.
REFERENCE 1 (bases 1 to 466)
AUTHORS Asamizu, E., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohyama, K.,
Nakamura, Y. and Tabata, S.
TITLE Generation of expressed sequence tags from low-CO2 and high-CO2
adapted cells of Chlamydomonas reinhardtii
JOURNAL DNA Res. 7 (5), 305-307 (2000)
PUBMED 11089912
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
Location/Qualifiers
1..466
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="C9"
/db_xref="taxon:3055"
/clone="LCL050a06_r"
/notes="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was constructed from cells cultured
in a carbon stress acclimatized condition in which carbon
dioxide concentration in the bubbling gas was changed from
5% to 0.04%"

ORIGIN
Alignment Scores: 2.5e-51 Length: 466
Pred. No.: 556.00 Matches: 109
Score: 85.9% Conservative: 1
Percent Similarity: 85.2% Mismatches: 0
Best Local Similarity: 16.1% Indels: 18
Query Match: 1
Gaps: 1
DB: 1

US-10-620-914-45 (1-648) x AV628989 (1-466)

Qy 1 MetGlySerGlyArgAspGlyArgProAlaSerTyrThrLysLysAsnPheSerLeuGlu 20
Db 82 ATGGGGTGGGTGCTGTCAGCGCGCGCTCGAGCTACACCAAGAAGAACTTCTCCCTGGAG 141

```



```
Qy 21 LysLeuLysLeuSerSerMetLysAspAspLeuThrValLeuArgHisMetTrrPheGly 40
Db 142 AAGCTCAGCTCAGCAGCAATGAAGATGACCTGACCGTTCTGCGCCATATGTGGTTCGGC 201
Qy 41 SerLysLysGlyAspAspHisAlaAlaArgLeuGluSerPheTyrGlyProGlnAlaAla 60
Db 202 AGCAAGAAGGGCGATGATCAGCTGCTCGCTGAGAGCTTCTACGGGCCCCAGGGCGCT 261
Qy 61 Alaphe----- 62
Db 262 GCCTATGATGCTTTCGGGTCCGCGTTCTCTCGGGTGCAGGCCCATGCTCGTCGAGTT 321
Qy 63 AlaAlaArgLeuAlaGluArgSerAsnLeuIleTrrValAspLeuGlyGlyThrGly 82
Db 322 GCTGCCCGCTGGCCGAGCGCTCGAACCTCATCTGGGTTGACCTGGGTGGTGGCAGCTGGG 381
Qy 83 GluAsnValAspMetMetAlaAspTyrIleAspLeuAlaLysPheLysSerIleTrrVal 102
Db 382 GAGAATGTCGATATGATGGCTGATTACATCGACCTGGCGAAGTTCAAGTCCATCTACGTG 441
Qy 103 ValAspLeuCysHisSerLeuCys 110
Db 442 GTCGACCTGTGCCACTCGCTGTGC 465
```

Search completed: March 14, 2006, 06:44:05  
Job time : 6860 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: March 14, 2006, 02:01:06 ; Search time 1035 Seconds  
(without alignments)  
4172.677 Million cell updates/sec

Title: US-10-620-914-45

Perfect score: 3463

Sequence: 1 MGSGRDPASYTKNFSLE.....RVNYSFFYMARKGAKDN 648

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DRV=xlh  
-Q=/abes/ABSSWEB spool/US10620914/runat\_13032006\_102038\_6713/app\_query.fasta.1  
-DB=N Geneseq -QWTF=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blom62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abes05h  
-USER=US10620914 @CGN 1.1 727 @runat\_13032006\_102038\_6713 -NCPU=6 -ICPU=3  
-NO WMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N\_Geneseq 21:\*  
1: Geneseqn1980s:\*  
2: Geneseqn1990s:\*  
3: Geneseqn2000s:\*  
4: Geneseqn2001as:\*  
5: Geneseqn2001bs:\*  
6: Geneseqn2002as:\*  
7: Geneseqn2002bs:\*  
8: Geneseqn2003as:\*  
9: Geneseqn2003bs:\*  
10: Geneseqn2003cs:\*  
11: Geneseqn2003ds:\*  
12: Geneseqn2004as:\*  
13: Geneseqn2004bs:\*  
14: Geneseqn2005s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3463	100.0	1947	14	Adw72747 Chlamydom
2	2838	82.0	5242	14	Adw72746 Chlamydom
3	891	25.7	2259	12	Adp98536 C. albica
4	850.5	24.6	2727	14	Adw72752 Neurospor

5	828	23.9	2540	13	ADR85469	Adr85469 Aspergill
6	804	23.2	3427	14	ADM72751	Adm72751 Neurospor
7	700	20.2	2929	13	ADR84882	Adr84882 Aspergill
8	700	20.2	8929	13	ADR84295	Adr84295 Aspergill
9	379.5	11.0	1252	8	ABS58051	Abes58051 Gene enco
10	379.5	11.0	1352	14	ADM72704	Adm72704 Rhodobact
11	329	9.5	1248	8	ABS58076	Abss58076 Mesorhizo
12	329	9.5	1248	14	ADM72725	Adm72725 Mesorhizo
13	326.5	9.4	1251	8	ABS58082	Abss58082 Agrobacte
14	326.5	9.4	1251	14	ADM72731	Adm72731 Agrobacte
15	292.5	8.4	1251	8	ABS58084	Abes58084 Sinorhizo
16	292.5	8.4	1251	14	ADM72735	Adm72735 Sinorhizo
17	138.5	4.0	8832	14	ACL64521	ACL64521 M. xanthu
18	128	3.7	4246	6	ABS78845	Abbs78845 E. coli C
19	128	3.7	4246	10	ADH80412	Adh80412 Escherich
20	126.5	3.7	990	14	ACL72740	Adl72740 M. xanthu
21	123	3.6	7119	13	ADT04702	Adt04702 House mou
22	121	3.5	639	10	ADC93973	Adc93973 E. faeciu
23	120	3.5	666	8	ABS58085	Abss58085 Sinorhizo
24	120	3.5	666	14	ADM72737	Adm72737 Sinorhizo
25	119	3.4	6912	13	ADD29805	Add29805 Human tum
26	119	3.4	6912	10	ADT04698	Adt04698 Human DNA
27	119	3.4	6912	14	ADZ49071	Adz49071 Insulin a
28	118.5	3.4	1155	3	AA065598	Aaa065598 Human imm
29	118.5	3.4	1155	3	AAC81011	Aac81011 Human B11
30	118.5	3.4	1155	4	AHH93714	Aah93714 Human pro
31	118.5	3.4	1155	4	AAI67211	Aai67211 B30SD iso
32	118.5	3.4	1155	4	AAI67211	Aai67211 B30SD iso
33	118.5	3.4	1155	4	AAI67211	Aai67211 B30SD iso
34	118.5	3.4	1155	4	AAH02779	Aah02779 Prostate
35	118.5	3.4	1155	4	AAH85028	Aah85028 Human pro
36	118.5	3.4	1155	5	ACA59615	Aca59615 Prostate
37	118.5	3.4	1155	6	AAS99857	Aas99857 Breast tu
38	118.5	3.4	1155	6	ACC95342	Acc95342 Prostate
39	118.5	3.4	1155	8	ADA11380	Adal1380 Human bre
40	118.5	3.4	1155	10	ADC15353	Adcl15353 Human bre
41	118.5	3.4	1155	10	ADB13823	Adb13823 Human pro
42	118.5	3.4	1155	10	ADG26239	Adg26239 Human pro
43	118.5	3.4	1185	8	ADA11414	Adal1414 Human CDN
44	118.5	3.4	1185	10	ADC15387	Adcl15387 Human bre
45	118.5	3.4	1383	8	ABZ79639	Abz79639 Mammaglob

ALIGNMENTS

RESULT 1	
ADW72747	
ID	ADW72747 standard; cDNA; 1947 BP.
XX	
AC	ADW72747;
XX	
DT	21-APR-2005 (first entry)
XX	
DE	Chlamydomonas reinhardtii Btal cDNA.
XX	
KW	Phospholipid synthesis; transgenic plant; fertilizer; ss; gene;
KW	betaine lipid.
XX	
OS	Chlamydomonas reinhardtii.
XX	
FH	Key
CDS	Location/Qualifiers
FT	1..1947
FT	/*tag= a
FT	/product= "Btal protein"
XX	
PN	WO2005009115-A2.
XX	
PD	03-FEB-2005.
XX	
PF	15-JUL-2004; 2004WO-US022789.
XX	
PR	16-JUL-2003; 2003US-00620914.
XX	

PA	(UNMS ) UNIV MICHIGAN STATE.	Db	421	ACCGCGACGCTCATCACCCTTCTCTACTCGCTCACGATGATCCACCGTTCCAAACGTC	480
XX					
PI	Benning C, Riekhof W;	Qy	161	IleAspGlnAlaCysSerTyrLeuSerGlnAspGlyLeuValGlyValAlaAlaAspPheTyr	180
XX					
DR	WPI; 2005-112975/12.	Db	481	ATCGACAGGCTTGCTGTAACCTTCTCCAAAGACGGCTGGTGGCGCTTCCGACTTCTAC	540
DR	P-PSDB; ADW72748.				
XX					
PT	New composition comprising a purified DNA having an oligonucleotide	Qy	181	ValSerGlyLysTyrAspLeuProLeuArgGlnMetProTyrSerArgArgPheTyr	200
PT	sequence encoding a protein, useful in producing Betaine lipid compounds	Db	541	GTGAGCGGCAAGTACGACCTGCCCTGCGCCAGATGCCCTGGTGGCGCGTTCCTCTGG	600
PT	e.g., Diacylglycerol-1-O-4'-(N,N,N'-trimethyl) homoserine (DGTS).				
XX					
PS	Claim 1; SEQ ID NO 44; 147pp; English.	Qy	201	ArgSerIlePheAspIleAspIleAspIleGlyProGluArgArgAlaTyrLeuGlu	220
PS		Db	601	CGATCGATCTTCGACATCGACCACTGCGCCCGGAGCGCGCGCTACCTGGAG	660
XX					
CC	The invention relates to a new composition comprising a purified DNA	Qy	221	GlnLysLeuGluArgValTyrGluGlnAsnThrGlnGlySerIleProTyrValProTyr	240
CC	having an oligonucleotide sequence appearing as ADW72747 (Chlamydomonas	Db	661	CAGAACTGGAGCGCTGTGGAGCAGAACACCCAGGGTTCGATCCCTACGTGGCGTG	720
CC	reinhardtii Btal cDNA) or ADW72752 (Neurospora crassa Btal coding region,				
CC	encoding Btal proteins which are DGTS (diacylglycerol-1-O-4'-(N,N,N'-	Qy	241	LeuArgAlaProTyrTyrValTyrIleGlyArgLeuProSerValGlyHisAlaLeuHis	260
CC	trimethyl)homoserine) synthetic enzymes. Btal (and BtaA and BtaB) enzymes	Db	721	CTGCGCGCCCTTACTACGTGTGGATTGGCGGCTGCCAGCGTTGGCCACGCGCTGCAC	780
CC	allow the replacement of phospholipids with non-phosphorus containing				
CC	lipids in the cell membrane. A transgenic plant expressing the enzymes	Qy	261	GluGluArgValGluArgProMetPheProThrPheLeuTyrThrGlnSerTyr	280
CC	would have a reduced need for phosphate-containing fertilizer. Also	Db	781	GAGGAGCGGTGGAGCGCGCGCCCATGTTCGCCGCCACCTTCCTGTACACGCGTGTGG	840
CC	included are an RNA transcribed from the purified DNA, antibodies	Qy	281	GluAspProGluProAspMetGluValMetGluIleAsnProLysAspThrValLeuThr	300
CC	produced from the protein, a vector comprising the DNA or nucleic acid, a	Db	841	GAGACCCCGAGCGCGATATGGAGGTGATGGAGATCAACCCCAAGACACGGTGTGACC	900
CC	host cell comprising the vector, transgenic plants comprising the vector,				
CC	and a protein translated from the RNA or encoded by the nucleic acid. The	Qy	301	LeuThrSerGlyGlyCysAsnAlaLeuAsnLeuValGlnGlyAlaGlyGlnValVal	320
CC	composition is useful in producing Betaine lipid compounds e.g.,	Db	901	CTGACTAGCGCGGCTGCAATGCCCTGTAACCTGCTGGTGAGGGGCGCGGAGGTGGTG	960
CC	Diacylglycerol-1-O-4'-(N,N,N'-trimethyl) homoserine (DGTS). The composition	Qy	321	SerValAspCysAsnProAlaGlnSerAlaLeuLeuGluLeuLysLysValAlaIleGln	340
CC	of the invention reduces the amount of phosphate fertilizer needed for	Db	961	TCGGTGGACTGCAACCCCGCGAGTCGGCGCTTCTGAGAGCTGAAGAGGTGGCGCATTCAG	1020
CC	the optimal growth of crop plants. The present sequence encodes a betaine	Qy	341	GlnLeuGluPheGluAspValTyrGlnLeuPheGlyGluGlyValHisProArgIleGlu	360
CC	lipid synthetic enzyme.	Db	1021	CAGCTGAGTTTGGAGACGTGTGGCAGCTGTTCTGGCGAGGGCGTGCACCCCGCATTTGAG	1080
XX					
SQ	Sequence 1947 BP; 365 A; 619 C; 613 G; 350 T; 0 U; 0 Other;	Qy	361	GluLeuTyrGluLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTyrSer	380
		Db	1081	GAGCTGTACGAGAAGAGCTGGCGCCTTCTGTCGCAACACAGCACCACTTCTGTGTCC	1140
		Qy	381	LysArgLeuTyrTyrPheGlnHisGlyLeuTyrTyrGlnGlyMetGlyLysLeuCys	400
		Db	1141	AAGCGCCTCTGGTACTTCCAGCAGCGCTGTACTACAGGGCGGCATGGGCAAGCTGTGC	1200
		Qy	401	TyrValLeuGlnCysLeuAlaValValLeuGlyLeuGlyLysThrValLysArgLeuAla	420
		Db	1201	TGGTGTCTGAGTGCCTGGCGCTGGTGTGGACTGGGCAAGACCGTCAAGCGCTCGCC	1260
		Qy	421	AsnAlaProThrMetGluGluGlnArgArgLeuTyrAspSerAsnMetLeuIleHisPhe	440
		Db	1261	AACGCGCCCAACTGGAGGAGCAGCGCGCTGTGGGACAGCAACATGCTCATCACTTC	1320
		Qy	441	ValLysAsnGlyProLysProLeuValTyrPheValLysPheValSerLeuValLeu	460
		Db	1321	GTGAAGAACCGGCGCAAGCGCTGGTGTGGCTGTTCGTCGTCGAGCTGGTGTCTC	1380
		Qy	461	PheAsnLysAlaValLeuTyrPheGlyGlyValProGlyLysGlnTyrAlaLeuIle	480
		Db	1381	TTCAACAGCGCGTGTGTGTGGCGGGCGGTGGCGGCAAGCAGTACGCGCTGATC	1440
		Qy	481	LysAlaAspGlyIleProIleGluAsnTyrIleAlaArgThrMetAspGlyValAlaGlu	500
		Db	1441	AAGCGGACGGCATCCCAATTGAGAACTACATCGCGCGCAACATGGAGCGCGTGGCGAG	1500
		Qy	501	AsnSerHisValArgLysGlnAsnTyrPheTyrTyrAsnCysLeuThrGlyLysPheLeu	520
		Db	1501	AACTCGCACGTGCGCAAGCAGAACTACTTCTACTACAACTGCCTCACCGCAAGTTCCTG	1560

## Alignment Scores:

Pred. No.: 0 Length: 1947  
 Score: 3463.00 Matches: 648  
 Percent Similarity: 100.0% Conservative: 0  
 Best Local Similarity: 100.0% Mismatches: 0  
 Query Match: 100.0% Indels: 0  
 DB: 14 Gaps: 0

US-10-620-914-45 (1-648) x ADW72747 (1-1947)

Qy	1	MetGlySerGlyArgAspGlyArgProAlaSerTyrThrLysLysAsnPheSerLeuGlu	20
Db	1	ATGGGGTGGGTGTGACGGCGCGCGCTGCGAGCTACACCAAGAGAACTTCTCCTCGAG	60
Qy	21	LysLeuLysLeuSerSerMetLysAspAspLeuThrValLeuArgHisMetTyrPheGly	40
Db	61	AAGCTCAAGCTCAGCAGCATGAAGGATGACCTGACCGTCTGCGCCCATATGTGTTGGC	120
Qy	41	SerLysLysGlyAspAspHisAlaAlaArgLeuGluSerPheTyrGlyProGlnAlaAla	60
Db	121	AGCAAGAAGGCGCATGATCAGCTGCTCGCTGCGAGAGCTTCTACGGGCGCCCAAGCGCGCT	180
Qy	61	AlaPheAlaAlaArgLeuAlaGluArgSerAsnLeuIleTyrValAspLeuGlyGly	80
Db	181	GCCTTTGTGCTCCCGCTGGCGGAGCGCTCGAACCTCATCTGGGTGACCTGGGTGGTGGC	240
Qy	81	ThrGlyGluAsnValAspMetMetAlaAspTyrIleAspLeuAlaLysPheLysSerIle	100
Db	241	ACTGGGGAGATGTCGATGATGATGCTGATACATCGACCTGGCGGAAGTTCAAGTCCATC	300
Qy	101	TyrValValAspLeuCysHisSerLeuCysGluValAlaLysLysLysAlaLysAlaLys	120
Db	301	TACGTGGTTCGACCTGTGCACTCGCTGTCGAGGTGGCGCAGAGAGAGCGGCAAG	360
Qy	121	GlyTyrLysAsnValGlnValValGluAlaAspAlaCysGlnPheAlaProProGluGly	140
Db	361	GGCTGGAGAGATGTCAGGTGTCGAGCGCGACCTTGGCCAAATTGGCCCTTGAAGGC	420
Qy	141	ThrAlaThrLeuIleThrPheSerTyrSerLeuThrMetIleProProPheHisAsnVal	160

QY 521 ArgAspAsnCysProThrThrTyrLeuArgGluAlaAlaPheAlaThrLeuLysSerGlyVal 540  
 Db 1561 CGCGACAACCTGCGCCACCTACCTGCGCGAGCGCGCTTCGCCACCTCAGAGTGGCGTG 1620  
 QY 541 ValAspAsnLeuThrValSerThrAsnPheMetClnuLysAlaAargThrTyr 560  
 Db 1621 GTGGACAACCTGACCGCTCCACCACTTCTTATGGAGAGGCTCAAGGCGCACCTAC 1680  
 QY 561 ThrLysValIleLeuMetAspHisValAspTyrLeuAspMetProValAlaAsnGluLeu 580  
 Db 1681 ACCAAGTGATTCGTGGACCACTGAGTGGCTGGATATGCCGTGGCCACAGAGCTG 1740  
 QY 581 AlaGluCysLeuAlaLysGlnValAlaProGlyGlyIleValIleTyrArgSerAlaSer 600  
 Db 1741 GCCGAGTGCTGCGCAAGCAGGTGGCGCGCGGCGCATCTCATCTGGCGCTCGCGCTCC 1800  
 QY 601 LeuSerProProTyrAlaGluLeuIleGlnLysAlaGlyPheAspValAlaArgCysIleAarg 620  
 Db 1801 CTGAGCGCGCCCTACGCGGAGCTGATCCAGAGCGCGGCTTCGACGTGGCGCTGCATCCGC 1860  
 QY 621 ArgAlaThrGlnGlyTyrMetAspArgValAsnMetTyrSerSerPheTyrMetAlaArg 640  
 Db 1861 CGCGCCACTCAGCGCTACATGGACCGGCTCAATGATGACAGCTCTCTTACATGGCGCGC 1920  
 QY 641 ArgLysGlyAlaLysLysAspAsn 648  
 Db 1921 CGGAAGGCGGCCAAGAAGGACAAC 1944

## RESULT 2

ADW72746  
 ID ADW72746 standard; DNA; 5242 BP.

XX AC ADW72746;

XX DT 21-APR-2005 (first entry)

XX DE Chlamydomonas reinhardtii Btal gene.

XX KW Phospholipid synthesis; transgenic plant; fertilizer; ds; gene;  
 XX KM betaine lipid.

XX OS Chlamydomonas reinhardtii.

XX PN WO2005009115-A2.

XX PD 03-FEB-2005.

XX PF 15-JUL-2004; 2004WO-US022789.

XX PR 16-JUL-2003; 2003US-00620914.

XX PX (UNMS ) UNIV MICHIGAN STATE.

XX PY Benning C, Riekhof W;

XX DR WPI; 2005-112975/12.

XX DR P-PSDB; ADW72748.

XX PT New composition comprising a purified DNA having an oligonucleotide  
 sequence encoding a protein, useful in producing Betaine lipid compounds  
 e.g., Diacylglycerol-1-O'-4'-(N,N,N-trimethyl) homoserine (DGTS).

XX PS Example 4; SEQ ID NO 43; 147pp; English.

XX CC The invention relates to a new composition comprising a purified DNA  
 having an oligonucleotide sequence appearing as ADW72747 (Chlamydomonas  
 reinhardtii Btal cDNA) or ADW72752 (Neurospora crassa Btal coding region,  
 encoding Btal proteins which are DGTS (diacylglycerol-1-O'-4'-(N,N,N-  
 trimethyl)homoserine) synthetic enzymes. Btal (and BtaA and BtaB) enzymes  
 allow the replacement of phospholipids with non-phosphorus containing  
 lipids in the cell membrane. A transgenic plant expressing the enzymes  
 would have a reduced need for phosphate-containing fertilizer. Also

CC included are an RNA transcribed from the purified DNA, antibodies  
 CC produced from the protein, a vector comprising the DNA or nucleic acid, a  
 CC host cell comprising the vector, transgenic plants comprising the vector,  
 CC and a protein translated from the RNA or encoded by the nucleic acid. The  
 CC composition is useful in producing Betaine lipid compounds e.g.,  
 CC Diacylglycerol-1-O'-4'-(N,N,N-trimethyl) homoserine (DGTS). The composition  
 CC of the invention reduces the amount of phosphate fertilizer needed for  
 CC the optimal growth of crop plants. The present sequence encodes a betaine  
 CC lipid synthetic enzyme.

XX SQ Sequence 5242 BP; 955 A; 1562 C; 1662 G; 1063 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 1.83e-296 Length: 5242  
 Score: 2838.00 Matches: 645  
 Percent Similarity: 45.6% Conservatives: 0  
 Best Local Similarity: 45.6% Mismatches: 3  
 Query Match: 82.0% Indels: 771  
 DB: 14 Gaps: 11

US-10-620-914-45 (1-648) x ADW72746 (1-5242)

QY 1 MetGlySerGlyArgAspGlyArgProAlaSerTyrThrLysLysAsnPheSerLeuGlu 20  
 Db 501 ATGGGGTGGGGTGGTGGCGCGCGCTGGCGAGCTACACCAAGAGAACTTCTCCCTGGAG 560  
 QY 21 LysLeuLysLeuSerSerMetLysAspAspLeuThrValLeuArgHisMetTyrPheGly 40  
 Db 561 AAGCTCAAGCTCAGCAGCATGAAGATGACCTGACCGCTTCTGCCCATATATGTGTTCCGC 620  
 QY 41 SerLysGlyAspAspHisAlaAlaArgLeuGluSerPheTyrGlyProGlnAlaAla 60  
 Db 621 AGCAAGAAGGCGCATGATCAGCTGCTGCGCTGGAGAGCTTCTACGGGCGCCAGCGCGCT 680  
 QY 61 AlaPhe----- 62  
 Db 681 GCCTG-TAAGTCGCGATTAACTTTATGCTCATTTAAGTGTTCGAGTTCGTCGCGCTG 739  
 QY 62 ----- 62  
 Db 740 TCTGCCCGCGCAGATGATGCTTTTCGGTTCGGTTCCTCTGCGGTGCGCAGGCCCATGCTC 799  
 QY 63 -----AlaAlaArgLeuAlaGluArgSerAsnLeuIleTyrValAspLeuGlyGly 79  
 Db 800 GTGCAGTGTGCTGCCCGCTGCGCGAGCGCTCGAACCTCATCTGGGTGACCTGGGTGGT 859  
 QY 80 GlyThrGly----- 82  
 Db 860 GGCACCTGGGGTGAGTGGCGGTTGTATCGGGAGGTGGGATAGCCCTGCGCGCTTGGCT 919  
 QY 83 -----GluAsnVal 85  
 Db 920 GCGCCTCGCGGCCATTATGCTGATCTGCGCTTTGCTTTGTCGCGCAGGAGAAATGTC 979  
 QY 86 AspMetMetAlaAspTyrIleAspLeuAlaLysPheLysSerIleTyrValValAspLeu 105  
 Db 980 GATATGATGCTGATTACATCGACCTGGCGAAGTTCAGTCCATCTACGTTGGTGCACCTG 1039  
 QY 106 CysHisSerLeuCysGluValAlaLysLysAlaLysAlaLysGlyTyrLysAsnVal 125  
 Db 1040 TGCACCTGCTGTCGAGGTGGCCAAAGAAGAGCGCAAGGCGTGGAAAGATGTC 1099  
 QY 126 GlnValValGluAlaAspAlaCysGlnPheAlaProGluGlyThrAlaThrLeuIle 145  
 Db 1100 CAGTCTGGAGGCGCGCTTGGCAATTTGCGCCCTGAGGGCACCGGACGCTCATC 1159  
 QY 146 ThrPheSerTyrSerLeuThrMet----- 153  
 Db 1160 ACCTTCTCTACTCGCTCACGAG-TGAGTTGCAACCGCTCGACTTGCATCGGAGGATC 1218  
 QY 154 -----IlePro 155  
 Db 1219 CATCCACCACCAACAGTTTCATCCCTCTCACCCCGCGCTTTTGTGTTGTCAGTGAATCCA 1278

156 ProPheHisValIleAspGlnAlaCysSerTyrIleuSerGlnAspGlyLeuValGly 175  
1279 CCGTTCACAAACGTCATCGACAGGCTGTCTGTACCTGTCCCAAGACGGCCCTGGTGGC 1338  
176 ValAlaAspPheTyrValSerGlyLysTyrAspLeuProLeuArgGlnMetProTrpSer 195  
1339 GTTGGCGACTTCTACGTGAGCGGCAAGTACGACCTGCCCTGTGGCAGATGCCCTGTGCG 1398  
196 ArgArgPhePheTrp 200  
1399 CGCGCTTCTCTGTGGCGTGAGTTACAGCGCGGCTACACATCTCTCAAGCAAGCGTA 1458  
200 200 200 200  
1459 GCTGAGCGGGCACAGAAAGCGCTTAGGGCCCGGGGTGCGCTTCTGGGTAGTGGCAAGGA 1518  
200 200 200  
1519 TATCGGCAAGGCCCTTCAGCTACCGAGGTACCATGGTGATCTGTCAAGCAGCTCATCAC 1578  
200 200  
1579 GGTGCTGACTGAGCAGCTGCTCGAGTCTGCGTCTGTATAGGCTTTTCAGCACATCACA 1638  
200 200  
1639 GCGAGTCTCCGAATGCGGTGCTGACAGGTATGCTCGGGACAGCACCATGCACCGT 1698  
200 200  
1699 GCGGCGTTCTTGACTGTCATCGCTCTCTCGGCGCTCCCTCGGTTCTGCTCAGACACGT 1758  
201 201  
1759 CTCGCTGCCCTGTGTCTGCTGCTGCTGTGTCAGATCGATCTTCGACATCGCAAC 1818  
209 IleAspIleGlyProGluArgAlaTyrIleuGlnLysLeuGluArgValTrpGlu 228  
1819 ATTGACATCGCCCGAGCGCGGCTTACTGGAGCAGAAGCTGGAGCGGTGTGGAG 1878  
229 GlnAsnThr 231  
1879 CAGAACACCCAGGTGTGTGCGCGTTGACACCTTGGCTGTGCAAGCCCAAGTAAGTTTG 1938  
231 231  
1939 CTCATCAGCCCTGCAGCACCGCGGTGCGCGGATATTATCATCGGCAGACCGTCCGCA 1998  
232 232  
1999 CTGTTGGTTAATCCACAGTCCGCTGCTTCTCTGCGTGTTCAGGGTTCGATCCCC 2058  
237 TyrValProTrpLeuArgAlaProTyrTyrValTrpIleGlyArgLeuProSerValGly 256  
2059 TACGTGCGTGGCTGCGCGGCCCTACTAGTGTGGATTGGCGCTTGCAGCGGTGGC 2118  
257 His 257  
2119 CG-TGAGTCGCTGCGCGGAGCCTGTCGCGATCAGGACACATGTTACGGGGGCGCTCA 2177  
257 257  
2178 GTGAAGCGTATCGTATCGTTGAGTCTGTCTCAAGGTACCAACAGCGCGTCCCTTG 2237  
257 257  
2238 CGCGTTGAAATTGTTGGCGCAAGCATGAGCGGTGCGCGGATGCCCAACTCGCGTTGAC 2297  
258 258  
2298 CCTGACAGCGCCCTGCAGGAGCGGTGAGCGCGGCCCATGTTCCCGGCCACCTTC 2357

275 LeuTyrThrGlnSerTrpGluAspProGluProAspMetGlu 288  
2358 CTGTACACGCACTCGTGGGAGGACCCCGAGCCGATATGAGGTGAGCGSCTCAGCCTGA 2417  
288 288  
2418 GGCAGATACGCACAAAGAAACGCTGTGTGTACCCGAGGCGGGATCAGCAGCTTCGG 2477  
288 288  
2478 AGCATTTGGCACCGGTCACTGTGCGACCAAGCAAGCAGCATAAGCACCAGCGGTT 2537  
288 288  
2538 CTGTGGCAATGTGCCACCGCAATGCTGSCCAGTGCAGCTTCGCAATTAGTGTATGCC 2597  
288 288  
2598 AGTATCAGCCTAGCTCAGCCTGCAAGCTGCTGTGTAGAAAGCAGCGATGTGGCACT 2657  
289 289  
2658 GAGCGGAACACTCTGCTGGAACCTGTGCTGCCATCTCTCTGCGCCCTTCCAGGTGAT 2717  
290 tGluIleAsnProLysAspThrValLeuThrLeuThrSerGlyGlyCysAsnAlaLeuAs 310  
2718 GGAGATCAACCCCAAGGACACGCTGTGACCTGACTAGCGCGGCTGCAATGCCCTGAA 2777  
310 nLeuLeuValGlnGlyAlaGly 317  
2778 CCTGCTGCTGAGGGGCGCGCCAGGTGAGATCATAGTGGTGTGTCACCTTCTGGTG 2837  
317 317  
2838 TCCTTACTCATGATGCCCGCTGCACGGAATGTTAGGAAGCGGTGGCATACATGTAGCT 2897  
318 318  
2898 GCAGAATGTCGTCAGCTGCTCTTCGCGCTGTGCTGTCAGGTGGTGTGCTGGTACATGCA 2957  
325 snProAlaGlnSerAlaLeuLeuGluLeuLysValAlaIleGlnGlnLeuGluPheG 345  
2958 ACCCGCGCAGTGGCGCTTCTGAGCTGAAGAAGGTGGCCATTCAGCAGCTGGAGTTG 3017  
345 lukAspValTrp 348  
3018 AGGACGTGTGGCAGGTAAAGGGGCTCTCTCATCTGGCGCTAGGACTGCACACGCTCTTTC 3077  
348 348  
3078 CTACTGTGATGCAACATGCGGACCTGGTCAAGCTTTCGCGCATTCGCACTGCGATGTG 3137  
348 348  
3138 TCCCCCAGGCTCAGCACATCTCTCTCGCGCTGCAACGGCAGACGCCCTTTTACACTCC 3197  
349 349  
3198 GCAACCCCTCTCTGTGTAATTCCTTCCACAGCTGTTCGCGAGGCGGTGCACCCGCGC 3257  
359 IleGluGluLeuTyrGluLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPhe 378  
3258 ATTGAGGAGCTGTACGAGAAGAGTGGCGCCCTTCTGTCGCAAAACAGCAGCAACTTC 3317  
379 TrpSerLysArgLeuTrpTyrPheGlnHisGlyLeuTyrTyrGlnGlyGlyMetGlyLys 398  
3318 TGTTCGAAGCGCTCTGTGTTCTCCAGCAGCGCTGTACTACAGGGCGGCGATGGGCAAG 3377  
399 LeuCysTrpValLeuGlnCysLeuAlaValValLeuGlyLeuGlyLysThrValLysArg 418  
3378 CTGTGCTGGTGTGTCAGTGCCTGGCGTGTGCTGGAGCTGGGCAAGACCGTCAAGCGC 3437  
419 LeuAlaAsnAlaProThrMetThrMetGluGlnArgLeuTrpAspSerAsnMetLeuIle 438

Db 3438 CTGCCAACCGCCCAATGGAGGACGCGCGCTGTGGGACAGCAACATGCTCATC 3497  
Qy |||||  
Db 439 HiePheVallyAsenGlyProLysProLeuValTriPLeuPheValLysPheValSerLeu 458  
Qy |||||  
Db 3498 CACTTCGTGAAGAAGCGGCCAGCCGCTGGTGGCTGTCTCAAGTTCTGTAGCGCTG 3557  
Qy |||||  
Db 459 ValLeuPheAsenLysAlaValLeuTriPLeuGlyValProGlyLysGlnTyrAla 478  
Db 3558 GTGCTCTTCAACAGGCGGTGTGTGTGGCGCGCGCTGCCGGCAAGCAGTACGCG 3617  
Qy |||||  
Db 479 LeuIleValAlaPheGlyIleProIleGluAsnTyrIleAlaArgThrMetAspGlyVal 498  
Db 3618 CTGATCAAGCGGACGCGATCCCATTTGAGAACTTACATCGCGGCACCATGGACGCGGTG 3677  
Qy |||||  
Db 499 AlaGluAsnSerHisValArgLysGlnAsnTyrPheTyrTyrAsnCysLeuThrGlyLys 518  
Db 3678 GCGGAGAACTCGCAGCTGCGCAGCAGAACTACTTCTACTACAACTGCTCAGCGCAG 3737  
Qy |||||  
Db 519 PheLeuArgPheAsnCysProThrTyrLeuArgGluAlaPheAlaThrLeuLysSer 538  
Db 3738 TTCTGCGCGACAACTGCCCACTACCTGCGGAGCGGCTTGGCACCTCAAGAGT 3797  
Qy |||||  
Db 539 GlyValValAspLeuThrValSerThrAsnPheMetGluLeuLysAlaArg 558  
Db 3798 GCGGTGTGTGACAACTGACCGCTCTCCACCACTTCTTCATGAGGAGGCTCAAAAGCGCG 3857  
Qy |||||  
Db 559 ThrTyrThrLys 562  
Db 3858 ACCTACCAAGGTGGGAGCCAGCGGAGAGTGTGAAAGCGGAGATGGCGGGCGTTG 3917  
Qy |||||  
Db 562 562  
Db 3918 TAGGTGCATGGCGAAACGCTTGTCTGCCAGGTGAAAGGGCAGAGCGAGGAAACGGAATT 3977  
Qy |||||  
Db 562 562  
Db 3978 TCAGAAATTCGGATGGTGGGAAAGCGGAAACGAAACAGAACGGAATAGAAAGCGCTGGCG 4037  
Qy |||||  
Db 562 562  
Db 4038 AGAACTGGTGGTGGCGTTTGGCATCCAGTCTGCCAGCGGCGCCGCTACCTACCTGCT 4097  
Qy |||||  
Db 563 563  
Db 4098 CAGGTCCGTCTAACCGTTACAAACCTCGCCCTCACCGCCCCCGCCCGCCACCGAGTG 4157  
Qy |||||  
Db 564 IleLeuMetAspHisValAspTyrLeuAspMetProValAlaAsnGluLeuAlaGluCys 583  
Db 4158 ATTCTGATGACACCACTGGGACTGGCTGGATATGCCCGTGCCCAACGAGCTGGCGGAGTGC 4217  
Qy |||||  
Db 584 LeuAlaLysGlnValAlaProGlyGlyIleValIleTyrArgSerAlaSerLeuSerPro 603  
Db 4218 CTGGCCAGCAGGTGGCGCGGGCGGATCGTATCTGGCGCTCCGCTCCCTCAGCCCG 4277  
Qy |||||  
Db 604 ProTyrAlaGluLeuIleGlnLysAlaGlyPheAspValArgCysIleArgAlaThr 623  
Db 4278 CCCTAGCCGAGCTGATCCAGAAAGCGGGCTTCGACGTGGCTGCATCCGCGCGCCACT 4337  
Qy |||||  
Db 624 GlnGlyTyrMetAspArgValAsnMetTyr 633  
Db 4338 CAGGGCTTACATGACCGCGTCAACATGTACAGGTGGCGTGTCTTGTTCGTGAGGTGTGAG 4397  
Qy |||||  
Db 633 633  
Db 4398 GATGTAGATGTCAAGTGGTGGCGGAGGAGCATGCGGTGCGATGGGAGGAGGTTCGT 4457  
Qy |||||  
Db 633 633  
Db 4458 TATTGATCCCGATGGGACCTGGTGAATGGGTGAGCGGCTACCGTATGTGCCAAGGAC 4517  
Qy |||||  
Db 633 633

Db 4518 GGGGTTGGCGCGCTCGGGCGGACAGCAAACTTTGAAAGGTTGTTTATAGATCAAGATA 4577  
Qy 633 633  
Db 4578 TGTTCGATTAACGGAAGCGACCTAGCGTCAGGGCGCAGACAGCGGTGGGACCGCGGAA 4637  
Qy 634 634  
Db 4638 TCTGCGTTGTCTTCGATGTATACCTGTATCCCAATCTTGTGGTTTCTCCAATGCGCAGC 4697  
Qy |||||  
Db 635 SerPheTyrMetAlaArgArgLysGlyAlaLysIleAspAsn 648  
Db 4698 TCCTTCTACATGTCGCCCGGAAAGGCGCCCAAGAGGACAAAC 4739  
RESULT 3  
ADP98536  
ID ADP98536 standard; DNA; 2259 BP.  
XX ADP98536;  
AC ADP98536;  
XX 23-SBP-2004 (first entry)  
DT XX  
XX C. albicans specific gene, orf6.3438, DNA sequence.  
DE XX  
XX Diploid fungal cell; allele; gene disruption cassette;  
KW promoter replacement fragment; antifungal; fungicide; gene therapy;  
KW infection; Candida albicans; gene; ds.  
XX  
OS Candida albicans.  
XX  
XX WO2004056965-A2.  
PN 08-JUL-2004.  
XX  
PD 19-DEC-2003; 2003WO-US040618.  
XX  
PF 19-DEC-2002; 2002US-0434832P.  
PR  
XX (ELIT-) ELITRA PHARM INC.  
XX (ELIT-) ELITRA CANADA LTD.  
XX  
XX Roemer T, Jiang B, Boone C, Bussey H;  
PI  
XX WPI; 2004-500296/47.  
DR P-PSDB; ADP98846.  
DR  
XX  
XX Constructing a strain of diploid fungal cells in which both alleles of a gene are modified comprises modifying the alleles of a gene in the fungal cells by recombination using a gene disruption cassette and a promoter replacement fragment.  
PT  
XX Claim 33; SEQ ID NO 6021; 163pp; English.  
PS  
XX The invention relates to a novel method for constructing a strain of diploid fungal cells in which both alleles of a gene are modified. The method comprises modifying the alleles of a gene in diploid fungal cells by recombination using a gene disruption cassette and a promoter replacement fragment. The invention further comprises: assembling a collection of diploid fungal cells each of which comprises modified alleles of a different gene; a strain of diploid fungal cells comprising modified alleles of a gene, where the first allele of the gene is inactivated by a gene disruption cassette comprising a nucleotide sequence encoding an expressible selectable marker; and the expression of the second allele of the gene is regulated by a heterologous promoter that is operably linked to the coding region of the second allele of the gene, and where the gene encodes the polypeptide mentioned above; a collection of diploid fungal strains comprising the diploid strains cited above, where substantially all the different genes that encode the above amino acid sequences are modified and are present in different diploid strains in the collection; a nucleic acid molecule comprising a nucleic acid sequence that is hybridizable to a target nucleotide sequence comprising any of the 310 nucleotide sequences listed in the

CC specification (ADP98516-ADP98825); identifying a gene that is essential  
 CC to the survival or growth of a fungus, that contributes to the virulence  
 CC and/or pathogenicity of a fungus, or that contributes to the resistance  
 CC of a diploid fungus to an antifungal agent; identifying an antifungal  
 CC agent that inhibits the growth of a diploid fungus, or a therapeutic  
 CC agent for treatment of a mammalian disease; correlating changes in the  
 CC levels of proteins or gene transcripts with the inhibition of growth or  
 CC proliferation of a diploid fungal cell; a purified or isolated nucleic  
 CC acid molecule comprising a nucleotide sequence encoding a gene product  
 CC required for proliferation of *Candida albicans*, where the gene product  
 CC consists of any of the above-mentioned amino acid sequences; a vector  
 CC comprising a promoter operably linked to the nucleic acid molecule cited  
 CC above; a host cell containing the vector; a purified or isolated  
 CC polypeptide comprising any of the 61 amino acid sequences given in the  
 CC specification (ADP96719-ADP96778); a fusion protein comprising a fragment  
 CC of a first polypeptide fused to a second polypeptide, the fragment  
 CC consisting of at least 6 consecutive residues of any of ADP98826-ADP99135  
 CC ; producing a polypeptide; identifying a compound which modulates the  
 CC activity of a gene product encoded by a nucleic acid comprising any of  
 CC ADP98516-ADP98825; eliciting an immune response in an animal; a strain of  
 CC *Candida albicans*, where a first allele of a gene comprising any of  
 CC ADP98516-ADP98825 is inactive and a second allele of the gene is under  
 CC the control of a heterologous promoter; identifying a compound or binding  
 CC partner that binds to the polypeptide comprising any of ADP98826-  
 CC ADP99135, or its fragment; identifying a compound having the ability to  
 CC inhibit growth or proliferation of *Candida albicans*; inhibiting growth or  
 CC proliferation of *Candida albicans* cells; manufacturing an antimycotic  
 CC compound; treating an infection of a subject by *Candida albicans*;  
 CC preventing or containing contamination of an object by *Candida albicans*,  
 CC or for preventing or inhibiting formation on a surface of a biofilm  
 CC comprising *Candida albicans*; a pharmaceutical composition comprising a  
 CC therapeutic amount of an agent which reduces the activity or level of a  
 CC gene product encoded by a nucleic acid comprising any of ADP98516-  
 CC ADP98825 in a pharmaceutical carrier; an antibody preparation which binds  
 CC the polypeptide; methods for evaluating a compound against a target gene  
 CC product encoded by any of ADP98516-ADP98825; identifying an antimycotic  
 CC compound; a computer or a computer readable medium that comprises at  
 CC least one of the nucleotide sequences mentioned in the specification or  
 CC at least one amino acid sequence selected from ADP98826-ADP99135; a  
 CC method assisted by a computer for identifying a putatively essential gene  
 CC of a fungus; and a protein array comprising proteins, where at least one  
 CC protein comprises an amino acid sequence or a portion of an amino acid  
 CC sequence selected from ADP98516-ADP98825. The novel methods and  
 CC compositions have fungicide activity. The compositions may be used in  
 CC gene therapy. The composition and methods are useful for drug screening  
 CC purposes or for diagnosing, preventing or treating infections associated  
 CC with *Candida albicans*. These may also be used for constructing strains  
 CC useful for identification and validation of gene products as effective  
 CC targets for therapeutic intervention, for identifying and validating gene  
 CC products as effective targets for therapeutic intervention, and for  
 CC collecting identified essential genes. This polynucleotide sequence  
 CC represents a *Candida albicans* fungal specific gene of the invention.  
 CC NOTE: This sequence was downloaded from an electronic sequence listing  
 CC provided on the WIPO website.

XX Sequence 2259 BP; 714 A; 359 C; 406 G; 780 T; 0 U; 0 Other;

# Alignment Scores:

Pred. No.:	1,12e-85	Length:	2259
Score:	891.00	Matches:	229
Percent Similarity:	46.6%	Conservative:	98
Best Local Similarity:	32.7%	Mismatches:	232
Query Match:	25.7%	Indels:	142
DB:	12	Gaps:	21

US-10-620-914-45 (1-648) x ADP98536 (1-2259)

Qy	51	LeuGluSerPheTyrGlyProGlnAlaAlaPhe-----	62
Db	208	TTAGAGTCATCTTCAAGAAATCAAGCTCATATTTATGATACACATAGAGAAATTTTGTG	267
Qy	63	-----AlaAlaArgLeuAlaGluArgSerAsnLeu	72

Db	268	AAAGTAGACAGAAATGCTTAGATTAGCTATTTCCTATTCACCAAAAAAAGATCTT	327
Qy	73	IleTyrValAspLeuGlyGlyThrGlyGluAsnValAspMetMetAlaAspTyrIle	92
Db	328	ATTGGATTGATATGTTGGTGAACCTGGTTCCAATATTCGAATTCATGGATGAATAGT	387
Qy	93	AspLeuAlaLys---PheLysSerIleTyrValValAspLeuCysHisSerLeuCysGlu	111
Db	388	AAAATATCTGAAACCTTTAAAGCTGTTTATTGTTGGTATCTTTCCCATCTTTGTGTGAA	447
Qy	112	ValAlaLysLysLysAlaLysGlyTyrPlyAsnValGlnValValGluAlaAsp	131
Db	448	GTTGCTAAGGCAAGATTGAGCCCATGAATGTCATGATGATGATGATGATGATGAT	507
Qy	132	AlaCysGlnPheAlaProGluGlyThrAlaThrLeuIleThrPheSerTyrSerLeu	151
Db	508	GCCTGTGATTTTACTATTGATGATGATGATGATGATGATGATGATGATGATGATGAT	567
Qy	152	ThrMetIleProProPheHisAsnValIleAspGlnAlaCysSerTyrLeuSerGlnAsp	171
Db	568	TCGATGATCCCAACTTTCATGCTGCTATCGATAATGCTGCTTCTTAATATGATATGAA	627
Qy	172	GlyLeuValGlyValAlaAspPheTyrValSer-----GlyLysTyr	185
Db	628	GGTATTATTATCCCACTGCTGATTTTGGTATTCAAAGCAGTCACACCTCAATGGGTCGATC	687
Qy	186	Asp-----LeuProLeuArgGlnMetProTyrSerArgAspPheTyrArg	201
Db	688	ATACTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT	747
Qy	202	SerIlePheAspIleAspIleAspIleGlyProGluArgAlaTyrLeuGluGln	221
Db	748	ATTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT	807
Qy	222	LysLeuGluArgValTyrGluGlnAsnThrGln-----GlySerIleProTyr	237
Db	808	AAATTTGGTACCGCTCAAACTCTTGAATTCATACAAAGGCTTTGGGTAATAATC	861
Qy	238	ValProTyrLeuArgAlaProTyrTyrValTyrIleGlyArgLeuProSerValGlyHis	257
Db	862	-----CCCTATTATTGATGATGATGATGATGATGATGATGATGATGATGATGAT	903
Qy	258	AlaLeuHisGluGlu-----ArgValGluArgProProMetPheProProThr	273
Db	904	ACCATTTTAGAAGATGAATTTGTTAGCCACTGAATCCCTTACCTTGTCTCCAACTACA	963
Qy	273	-----	273
Db	964	ACTCCAATCGTAATCAACTGAAGATATTCCAATTTCTAAAGTCATGAAGTCGCTTTA	1023
Qy	273	-----	273
Db	1024	ATCACTTGCACAAAAAATTTACCTTACCCTATCAATGACTATCAAAAGGAATATTGGAGA	1083
Qy	274	-----PheLeuTyrThr	277
Db	1084	GTCTACTATGATGAATGAATCCATGTTATGAACCAATTTAAAAACCAATATCATTTATGCT	1143
Qy	278	GlnSerTyrGluAspProGluProAspMetGluValMetGluIleAsnProLysAspThr	297
Db	1144	TTCACTTGGGAAGATCCCTCGTGAAGATCATATAAATTTTGAATTTTACAGTGATGATACT	1203
Qy	298	ValLeuThrLeuThrSerGlyGlyCysAsnAlaLeuAsnLeu-----LeuValGlnGly	315
Db	1204	GTTTGGCTATTACTTCAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT	1263
Qy	316	AlaGlyGlnValValSerValAspCysAsnProAlaGlnSerAlaLeuLeuGluLeuLys	335
Db	1364	CCAAAAAGATTCATGCTGTTGATCTTAAATCCATGTCATAAACCATTATTATGAAATGAAA	1323
Qy	336	LysValAlaIleGlnGlnLeuGluPheGluAspValTyrGlnLeuPheGlyGluGlyVal	355
Db	1324	TTGGCTAGTTTATGATGCTCTTCTCAAGAAACAAATTTGGTCAATGTTTGGTGAAGGTAA	1383



Qy 356 HisProArgIleGluGluLeuTyrGluLeuLysLeuAlaProPheLeuSerGlnThrSer 375  
 Db 1384 ATCGAAATTTCAATGATCTTTGATTGATCTTTGGCAGCGCACATGTCTTCTAATGCC 1443  
 Qy 376 HisAsnPheTrpSerLysArg-----LeuTrpTyrPheGlnHisGlyLeuTyrTyrGln 393  
 Db 1444 TTCCAAATCTGGATGAAGTAAAGACCTAAACCTTTTCTGGTAAAGTCTTTAT----- 1497  
 Qy 394 GlyGlyMetGlyLysLeuCysTrpValLeuGlnCys-----LeuAlaValValLeu 410  
 Db 1498 --GATACCTGGTTTCTAGATGGCATTAAAGATATCAAGATATGTTTTCAGAGTTGC 1554  
 Qy 411 GlyLeuGlyLysThrValLysArgLeuAlaAsnAlaProThrMetGluGlnGlnArgArg 430  
 Db 1555 GGTGTTAGTAATACGTTGAAGAACTTTGTCTGCAACCAACCACTGGGAAGCAATTCGAGA 1614  
 Qy 431 LeuTrpAspSerAsnMetLeuIleHisPheValLysAsnGlyProLysProLeuValTrp 450  
 Db 1615 ATTTGGAAATGAACATTTG-----AAACCAACT----- 1641  
 Qy 451 LeuPheValLysPheVal----SerLeuValLeuPheAsnLysAlaValLeuTrpPheGly 469  
 Db 1642 TTATTTAATCCTGTGGTGGTTCATTACTTGTGGTAAATCAATGTTTTCAGAAAGCT 1701  
 Qy 470 GlyGlyValProGlyLysGlnTyrAlaLeuIleLysAlaAspGlyIleProIleGluAsn 489  
 Db 1702 TTGGAGTTCAGCTAATCAAGCAGCTTTAATG-----GGACCATCTGTTATTAAA 1752  
 Qy 490 TyrIleAlaArgThrMetAspGlyValAlaGluAsnSerHisValArgLysGlnAsnTyr 509  
 Db 1753 TATGTTGTTGATCTTTGGACCAATTCATTAAGAGATCGATGATTTCAAAATGATACTAT 1812  
 Qy 510 PheTyrTyrAsnCysLeuThrGlyLysPheLeuArgAspAsnCysProThrTyrLeuArg 529  
 Db 1813 TTCTACTATTTATGATGATGGGAGATACACCAAAACAAATTTGTCAGATTTATTAAT 1872  
 Qy 530 GluAlaAlaPheAlaThrLeuLys-----SerGlyValVal 541  
 Db 1873 ACAAAAGGTTTCAACAGATTATCTAGCAGCTGCTACTGCCAGCGGATCATCTCCAAT 1932  
 Qy 542 AspAsnLeuThrValSerThrAsnPhePheMetGlu-----GluLeuLysAlaArg 558  
 Db 1933 GACAACTTGAATCCACACTGACACTTTAAATGAAAGTGTTCGTAGATTTAAAGAAAA 1992  
 Qy 559 ThrTyrThrLysValIleLeuMetAspHisValAspTrpLeuAsp-----MetPro 575  
 Db 1993 TCAATCACTATTCATGATCATGATCATGATGATGATGATGATGATGATGATGATGAT 2052  
 Qy 576 ValAlaAsnGluLeuAlaGluCysLeuAlaLysGlnValAlaProGlyGlyIleValIle 595  
 Db 2053 GCTATTAAATGAATTAATCT--GCTTTGAAAGATGCTGCTGCCAGCGGATGATGATTA 2109  
 Qy 596 TrpArgSerLysLeuSerProProTyrAlaGluLeuIleGlnLysAlaGlyPhe--- 614  
 Db 2110 CTTAGATTCAGCAAGTCAAAAGCCCTTGTACTTGAAGCAATTCAGAACTTTGGGATCCAA 2169  
 Qy 615 AspValArgCysIleArgArgAlaThrGlnGlyTyrMetAspArgValAsnMetTyrSer 634  
 Db 2170 GAAGAGAAATGTTGTTGCTCAACCTGTTCAAGTATGACAGAGTTTATATGATGATGCC 2229  
 Qy 635 Ser 635  
 Db 2230 AAT 2232  
 RESULT 4  
 ID ADW72752  
 XX ADW72752 standard; DNA; 2727 BP.  
 AC ADW72752;  
 XX  
 DT 21-APR-2005 (first entry)  
 XX

DE Neurospora crassa Btal coding region.  
 XX Phospholipid synthesis; transgenic plant; fertilizer; ss; gene;  
 KW betaine lipid.  
 XX Neurospora crassa.  
 XX Key Location/Qualifiers  
 FT CDS 1..2727  
 FT /\*tag= a  
 FT /product= "Btal"  
 XX  
 PN WO2005009115-A2.  
 XX 03-FEB-2005.  
 XX 15-JUL-2004; 2004WO-US022789.  
 XX 16-JUL-2003; 2003US-00620914.  
 XX (UNMS ) UNIV MICHIGAN STATE.  
 PA Benning C, Riekhof W;  
 XX WPI; 2005-112975/12.  
 DR P-PSDB; ADW72753.  
 XX  
 PT New composition comprising a purified DNA having an oligonucleotide  
 PT sequence encoding a protein, useful in producing Betaine lipid compounds  
 PT e.g., Diacylglycerol-O-4'-(N,N,N-trimethyl) homoserine (DGTS).  
 XX Claim 1; SEQ ID NO 49; 147pp; English.  
 CC  
 CC The invention relates to a new composition comprising a purified DNA  
 CC having an oligonucleotide sequence appearing as ADW72747 (Chlamydomonas  
 CC reinhardtii Btal cDNA) or ADW72752 (Neurospora crassa Btal coding region,  
 CC encoding Btal proteins which are DGTS (diacylglycerol-O-4'-(N,N,N-trimethyl)homoserine) synthetic enzymes. Btal (and BtaA and BtaB) enzymes  
 CC allow the replacement of phospholipids with non-phosphorus containing  
 CC lipids in the cell membrane. A transgenic plant expressing the enzymes  
 CC would have a reduced need for phosphate-containing fertilizer. Also  
 CC included are an RNA transcribed from the purified DNA, antibodies  
 CC produced from the protein, a vector comprising the DNA or nucleic acid, a  
 CC host cell comprising the vector, transgenic plants comprising the vector,  
 CC and a protein translated from the RNA or encoded by the nucleic acid. The  
 CC composition is useful in producing Betaine lipid compounds e.g.,  
 CC diacylglycerol-O-4'-(N,N,N-trimethyl) homoserine (DGTS). The composition  
 CC of the invention reduces the amount of phosphate fertilizer needed for  
 CC the optimal growth of crop plants. The present sequence encodes a betaine  
 CC lipid synthetic enzyme.  
 XX Sequence 2727 BP; 616 A; 766 C; 713 G; 632 T; 0 U; 0 Other;  
 SQ  
 Alignment Scores:  
 Pred. No.: 3,73e-81 Length: 2727  
 Score: 850.50 Matches: 239  
 Percent Similarity: 44.4% Conservative: 109  
 Best Local Similarity: 30.5% Mismatches: 240  
 Query Match: 24.6% Indels: 198  
 DB: 14 Gaps: 23  
 US-10-620-914-45 (1-648) x ADW72752 (1-2727)  
 Qy 19 LeuGluLysLeuLysLeuSerSerMetLysAspLeuThrValLeuArgHisMet--- 37  
 Db 449 TTGGAGAGCTGGAGATG---GGCTTGAANAAGACAAGGAATGGAGACATGTGTAA 505  
 Qy 38 ---TrpPheGlySerLysGlyAspAspHisAlaAlaArgLeuGluSerPheTyrGly 56  
 Db 506 CCGTGGCCGGGACAG-----GGA 523  
 Qy 57 ProGlnAlaAlaPheAlaAlaArgLeuAlaGluArgSerAsnLeuIleTrpValAsp 76

524	CCG	-----GGACAAGGAGAAACCG-ATATCGGTTAGAT	555
77	LeuGlyGlyThrGlyGluAsnValAspMetMetAlaAspTyrIleAspLeuAlaLys	96	
556	GTCCGTGGGGGCACAGCTGGATATCGAAGCCATGCCAAGTTTGTCAACGTCCTGAA	615	
97	---PheLysSerIleTyrValValAspLeuCyHisSerLeuCySGluValAlaLysLys	115	
616	TTCTTCAAGACTGTTTACCTAGTAGCACTTTTCGCCGTCACTTTGTGAAGTGGCTAGGAAG	675	
116	LysAlaLysAlaLysGlyTTrpLysAsnValGlnValValGluAlaAspAlaCySGlnPhe	135	
676	CGGTTTCCCAAGCTGGGTGGAGATGTGAGAGTTATCTGCACGGATGCTCGCAAGTTT	735	
136	Ala-----ProProGluGlyThr-----	141	
736	AGGCTTGAGGATATGAGGATGTTGACGAAGGAGAGTCTGGCTCTGGAGATTCTTCGCCT	795	
142	-----	145	
796	TCCTTGTGGGTTCGTGGGGGAGACGAAGCGGGAGCACATGCGGGAGCTGAGTTGATC	855	
146	ThrPheSerTyrSerLeuThrMetIleProPheHisAsnValIleAspGlnAlaCyS	165	
856	ACCATGCTTTATAGCCTTTCGATGATCCGGATTATTTCTCGAATATPCGATTCGCTCGAG	915	
166	SerTyrLeuSerGlnAspGlyLeuValGlyValAlaAspPheTyrValSerGlyLysTyr	185	
916	TCTCTGTAGCACCTCAGCGCTTGATTCGGTCTGTGACATTTTACGCCCACTCGAAAGTC	975	
186	Asp-----	196	
976	GACTTCACATTCGCCAACTACCGGGTGTCTTATGAACCGACACGTTGGCTATTTCGCG	1035	
197	ArgPhePheTyrArgSerIlePheAspIleAspAsnIleAspIleGlyProGluAArgArg	216	
1036	CGGAACCTTCGGCGCTCGTGGTTCGATGCTGCACAGGGTGTCTCTTGAGCCAGCTCGTCGA	1095	
217	AlaTyrLeuGluGlnLysLeuGluAArgValTrp-----GluGlnAsnThrGln	232	
1096	GATTATCTCGAGTACAGGTCGGGACTGTCTCTGACCGTCAACGCCCGCAACACACTTTC	1155	
233	GlySerIleProTyrValProTrpLeuArgAlaPro-----TyrTyrVal---	247	
1156	GGAGCAATTCCTACTACATCTGTTGGG-ATGCCTCAAGAAGCCCTTTTCTACGTGAG	1214	
248	-----	253	
1215	TCTACCACGAAATTGTGGAACACATCGATGCTATTTCGACAGAGTCCCAAGATCATC	1274	
253	-----	253	
1275	ACCCCGTCTAGTGGGCAAAATCTTCTCTAGCAAAATCGGCTACGCTTTCGAGTCGG	1334	
254	-----SerValGlyHisAlaLeu-----HisGluGluAArgValGlu	265	
1335	CCGCACAGCGCGGAGATGCGCTCAAAGCCCTTCAATACGCCCATCGAACAATCTCGGC	1394	
266	ArgPro-ProMet-----	269	
1395	CAACCTACCTCTCCGTCCTTCTTACCAAAATCACCACTGGAGGATCTACTACGACGA	1454	
270	-----PheProProThrPheLeuTyrThrGlnSerTrpGluAs	282	
1455	TCAATCCCGAAGCACACCCAGTTTCAATGACGAGTACATCTACCCCTTTTACCTGGGAAGA	1514	
282	pProGluProAspMetGluValMetGluIleAsnProLysAspThrValLeuThrLeuTh	302	
1515	CTCGCGGTGCACAGAAATCTCTTAACCTCGGGCCCGACGACGTGCTCTAGCCATC	1574	
302	rSerGlyGlyCyAsnAlaLeuAsnLeuLeuValGlnGlyGlnValValSerVa	322	
1575	CAGGCGCGGCACAACATCTTCTTCTACTCTGATGCAGAGTCCCGCTCGCGTGACAGCAT	1634	

322 lAspCysAsnProAlaGlnSerAlaLeuLeuGluLeuLysValaIleGlnGlnLe 342  
 1635 CGACCTAAACCCAGCCAAACACCTGCTTGAATCAAAAGTCGCTCTTTTACGACTCT  
 342 uGluPheGluAspValTrpGlnLeuPheGlyGluGlyValHisProArgIleGluGluLe 362  
 1695 GGATTTACCCGAGCTGTGGAAGATCTTCGGTGAAGGCAACACCCGACCTTCGCTCACT 1754  
 362 uTyrGluLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrpSerLysAr 382  
 1755 GCTCATCTCCAAACTCTCCCTCACCTCTCCGCGCGCGCTTTTCAATATCTGGGTATCCAA 1814  
 382 gLeuTrpTrpPheGln-----HisGlyLeuTyrTyrGlnGlyGlyMetGlyLy 398  
 1815 TGGCGACATATTACCGACCTCGGGGCGCGGTCTCTATGATACCGCGCGCTCCCGATA 1874  
 398 sLeuCysTrpValLeuGlnCysLeuAlaValValLeuGlyLeuGlyLysThrValLysAr 418  
 1875 CGCTATCCGTTCTTCGCTGGATTTCACACACTTCTTCTTCGCGCTCCGCGTCCGTCG 1934  
 418 gLeuAlaAsnAlaProThrMetGluGlnArgLeuTrpAspSerAsnMetLeuIl 438  
 1935 ACTTCTCTACTCCCAACCTCGAAGGGCAACGTTCCATCTACCAACCAAGATT----- 1989  
 438 eHisPheValLysAsnGlyProLysProLeuValTrpLeuPheValLysPheValSer-- 457  
 1990 -----CGTCCCTGT-----CTGCTCAACCGCTTCGTCAACGG 2021  
 458 -LeuValLeuPheAsnLysAlaValLeuTrpPheGlyGlyGlyValProGlyLysGlnTy 477  
 2022 CCTGGTCTCAGCTCCGACGCTTCTCTGTCGGCTTTGGGCGTGCCCAAGAATCAAAGT 2081  
 477 rAlaLeuIleLysAlaAsp----- 483  
 2082 GGCTATGATCGAAGCCGACTACCAACCGCGCTTCTATCTCTCTCCACACCCCGCAGCAG 2141  
 484 -----GlyIleProIleGluAsnTyrIleAlaArgThrMetAspGl 497  
 2142 CAAAGAAAACCAAGCGCGCCGCAATCTCTCCACTACACACCTCCACCCCTGTATCC 2201  
 497 yValAlaGluAsnSerHisValArgLysGlnAsnTyrPheTyrTyrAsnCysLeuThrGl 517  
 2202 CGTTCCTCCACCTCCCACTTCGCTCGGACAAACCTTTACTACTCTGCTGTGTCCTGGG 2261  
 517 yLysPheLeuArgAspAsnCysProThrTyrLeuArgGluAlaAlaPheAlaThrLeuLy 537  
 2262 ACATACACAGCAGTGCCATCCGNTTACTTTCCTCGCGCCCACTCTATACTCAG 2321  
 537 sSer---GlyValValAspAsnLeuThrValSerThrAsnPhePheMetGluGluLeu-- 555  
 2322 CGCTCTCGAGCGCTTTCAGCGCTTACGCATCCACACCGATCAAATAACAGGAGGTGTGGC 2381  
 556 -----LysAlaArgThrTyrThrLysValIleLeuMetAspHisValAspTrpLeuAs 573  
 2382 TAGGTTTCACCGGGTACTTTGACAGTAGCGGTGGTATGATGATGATGATGATGATGATG 2441  
 573 pMetPro-----ValAlaAsnGluLeuAlaGluCy 583  
 2442 TCCGCTTCGCTGAGGAGGAAAGGAGGCGAGCGGAGGAGCAAGTGAGAG 2501  
 583 sLeuAlaLysGlnValAlaProGlyGlyIleValIleTrpArgSerAlaSerLeuSerPr 603  
 2502 GTTCAATCGGCGGTGGAAGGTGGGTGGAAGGTGTTGTTGAGGAGCGCGGAGTGGAGCC 2561  
 603 oProTyrAlaGluLeuIleGlnLysAlaGlyPheAspValArgCysIle----- 619  
 2562 GTGGTATGTAGGGTTTTTTTGGAGGAAGGGTTTGGACGAGGAGGGTGGTTCGTGTA 2621  
 620 ----ArgArgAlaThrGlnGlyTyrMetAspArgValAsnMetTyrSerSerPheTyrMe 638  
 2622 ATCCGGAAGGGGACACGAGGTGATATGACAGCGTGAATATGATGATGATGATGATGATG 2681

QY 638 talaArgArg 641  
 Db 2682 CTTGGAAG 2691

RESULT 5  
 ADR85469  
 ID ADR85469 standard; DNA; 2540 BP.  
 XX ADR85469;  
 DT 04-NOV-2004 (first entry)  
 XX Aspergillus fumigatus essential gene open reading frame #106.  
 DE Fungicide; Aspergillus fumigatus infection; Farmer's lung disease;  
 KW drug screening; ds.  
 KW Aspergillus fumigatus.  
 OS  
 XX W02004067709-A2.  
 PN 12-AUG-2004.  
 XX 16-JAN-2004; 2004WO-US001099.  
 XX 17-JAN-2003; 2003US-0441281P.  
 PR 13-JUN-2003; 2003US-0478196P.  
 XX (ELIT-) ELITRA PHARM INC.  
 PA (ELIT-) ELITRA CANADA LTD.  
 XX  
 PI Jiang B, Hu W, Lemieux S, Roemer T;  
 DR WPI; 2004-594200/57.  
 DR P-PSDB; ADR86056.  
 XX  
 PT New purified or isolated Aspergillus fumigatus nucleic acid molecule  
 PT encoding a gene product, useful for diagnosing and/or treating invasive  
 PT fungal infections, such as Farmer's lung disease.  
 XX  
 PS Claim 2; SEQ ID NO 2106; 164pp; English.  
 XX  
 CC The present invention relates to Aspergillus fumigatus genes that are  
 CC essential and are potential targets for drug screening. The methods and  
 CC compositions of the present invention are useful for diagnosing and/or  
 CC treating invasive Aspergillus fumigatus infection, including the allergic  
 CC forms of the disease, such as Farmer's lung disease. They can also be  
 CC used in various drug discovery purposes, such as expression of the  
 CC recombinant protein, hybridization assay and construction of nucleic acid  
 CC arrays. The present sequence represents an Aspergillus fumigatus  
 CC essential gene open reading frame, used during diagnosis and drug  
 CC development in the invention. These genes share a high degree of sequence  
 CC conservation with known essential genes of candida albicans. The sequence  
 CC data for this patent is not represented in the printed specification, but  
 CC was obtained in electronic format from WIPO.  
 XX  
 SQ Sequence 2540 BP; 589 A; 704 C; 626 G; 621 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 9.25e-79 Length: 2540  
 Score: 828.00 Matches: 234  
 Percent Similarity: 42.8% Conservative: 103  
 Best Local Similarity: 29.7% Mismatches: 267  
 Query Match: 23.9% Indels: 185  
 DB: 13 Gaps: 19

US-10-620-914-45 (1-648) x ADR85469 (1-2540)

QY 18 SerLeuGluLysLysLeuSerSerMetLysAspAspLeuThrValLeuArgHisMet 37  
 Db 137 AGTTGACATACATAGCGGTATCTTACCTACCTCAGATCATCTATGATCGTTCCTTA 196  
 QY 38 TrpPheGlySerLysLysGlyAspAspHisAlaAlaArgLeuGluSerPheTyrGlyPro 57

Db 197 AGCCC-CACAAGAAAGGTCTCAATGCCAGCAAGATGCGCTAGAAAGTCTTCAAGACT 255  
 QY GlnAlaAlaAlaPhe- 62  
 Db 256 CAGGCTGGTGTATATGATGCTCCACTCGCAAGCGTCTTCTGTGTGGCGTGAGGATATGTTG 315  
 QY 63 - - - - -AlaAlaArgLeu- - - - -AlaGluArgSerAsn 71  
 Db 316 GGGCTTGTTCGCCCGCAATTGAAGTACAAGGTTGAGAACAAAGAACTCCAGGCTGGAAG 375  
 QY 72 LeuLeuTrpValAspLeuGlyGlyThrGlyGluAsnValAspMetMetAlaAspTyr 91  
 Db 376 GCAATCTGGGTGATATGGCGAGGACTGGCTACAACATTGAGGCCATGGCTCTCTTC 435  
 QY 92 IleAspLeuAlaLysPheLysSer- - -IleTyrValValAspLeuCysHisSerLeuCys 110  
 Db 436 CTACCAAGTACCAATTCTTCTCGCAGCTGTACTTGTGACCTGTCTCCCTCTCTTTC 495  
 QY 111 GluValAlaLysLysLysAlaLysGlyTyrLysAsnValGlnValValGluAla 130  
 Db 496 GAGGTTGCCCGTCAGCGATTTCGAGCGCTAGGCTGGAAGAAATGTACGGTCTGTCTGCAA 555  
 QY 131 AspAlaCysGlnPhe- - - - -AlaProPro 138  
 Db 556 GATGCGCGTCTTTCGCTCTCCCATGAGAAAGTCGACCCCGTCCCGCAGCTCCATCA 615  
 QY 139 GluGlyThrAlaThrLeuIleThrPheSerTyrSerLeuThrMetIleProProPheHis 158  
 Db 616 ACTGCAGCGCAGATCTTATCACCAGCTACAGCTTGTCCATGATCTCTGACTACTAC 675  
 QY 159 AsnValIleAspGlnAlaCysSerTyrLeuSerGlnAspGlyLeuValGlyValAlaAsp 178  
 Db 676 AGTGTGCTCGACTCCCTGACCGACCTCTTGGGGCGTCTGGTATCCTTGTGTCTCGAC 735  
 QY 179 PheTyrValSerGlyLysTyrAspLeuProLeu- - - - - 189  
 Db 736 TTTATGTGCAAGTATCTGACGTTTCGCCCGCAACTACATTTGGTGGTGGTCTTCAAT 795  
 QY 190 ArgGlnMetProTyrSerArgPhePheTyrArgSerIlePheAspIleAspAsnIle 209  
 Db 796 CGCCAGCTTAATGGCTGGGTCTGCTGCTTTGGCGGCGCTGCTTCATGCGGACCGTCTC 855  
 QY 210 AspIleGlyProGluArgAlaTyrLeuGluGlnLysLeuGluArgValTyrGluGln 229  
 Db 856 AGTCTGGAAGCAGCAGCAGAGACTATCTGGAATACCGCTTTGGAACCGCTCATCTGCC 915  
 QY 230 AsnThrGln- - - - -GlySerIleProTyrValProTyrLeuArgAlaProTyr 245  
 Db 916 AGCGAGCGCAATTATCTTCTCGCGCGCAAT- - - - -CCTTAC 951  
 QY 246 TyrValTrpIleGlyArg- - - - - 251  
 Db 952 TACATCTTCTGGAGCTCACAGAATCTTGATCAAAATCTCTCCGGTCAGGAGACAATT 1011  
 QY 251 - - - - - 251  
 Db 1012 GAAATGTTGGACGCTCTTTTACTGAAATCTCTTATCTTCTCCGCCCAATCACCAAG 1071  
 QY 251 - - - - - 251  
 Db 1072 GAGATGGAGAAGCAATGGAGAGAACGCTCAAGAAATTCGTTCCAAAGCTACGAGTCG 1131  
 QY 252 - - - - -LeuProSerValGlyHisAlaLeuHisGlu 261  
 Db 1132 GCGGTTATCACTGAGCGCGAACCCTCCCGCTTCCCTCATCTCTTACCAGAAATCACAC 1191  
 QY 262 GluArgVal- - - - -GluArgProProMetPheProProThrPheLeu 275  
 Db 1192 TCGCGCATCTTCTACAATGATCTGTGCGCAACACACGACGCTTCAAGAAATGAATACATT 1251  
 QY 276 TyrThrGlnSerTyrGluAspProGluProAspMetGluValMetGluLeuAsnProLys 295

```
Db 1252 TATGCTTCACTGGGAAGATCCTCGTGTGGATCATCTCTTCGCATCAAGCGGAT 1311
Qy 296 AspThrValLeuThrLeuThrSerGlyGlyCysAsnAlaLeuAsnLeuLeuValGlnGly 315
Db 1312 GATGTTATCTTGGCCATTTACAGCGCGGAGACAATATTTGGATTACCTTCAGAAAGAT 1371
Qy 316 AlaGlyGlnValValSerValAspCysAsnProAlaGlnSerAlaLeuLeuGluLeuLys 335
Db 1372 CACGAGAGATTTCATGCGAGTCGAGTCAATCTTCAACAGAGATCATTTGCTTGAACCTCAAG 1431
Qy 336 LysValAlaIleGlnGlnLeuGluPheGluAspValTrpGlnLeuPheGlyGluGlyVal 355
Db 1432 GTTGCTAGTTTCATGCGCCCTTGTCTCGCATGCTCTGGAAGATCTTTGGCGAGGAAAA 1491
Qy 356 HisProArgIleGluGluLeuThrGlyGlyLeuAlaProPheLeuSerGlnThrSer 375
Db 1492 CACCCAGAATTCAGGGAACCTCTCATTTCTCGTCTCAGCGCTCACCTCTCCAGCCAGGCA 1551
Qy 376 HisAsnPheTrpSerLysArgLeuTrpTyrPheGln-----HisGlyLeuTyr 391
Db 1552 TTCAGTACTGGCTTGAGCACACTCACATTTTCACTCAAAATATGGCAAGACTTTAT 1611
Qy 392 TyrGlnGlyMetGlyLysLeuCysTrpValLeuGlnCysLeuAlaValValLeuGly 411
Db 1612 GAAACGGTGGCTCGGCCACGCCATCAAGATGTTGGGTACCTGTTCAAGGTGTTTGGC 1671
Qy 412 LeuGlyLysThrValLysArgLeuAlaAsnAlaProThrMetGluGlnArgArgLeu 431
Db 1672 CTTGAGGGTCAAGTGAAGAAGCTCTCGAGCGCGAGACTCTTGTGAGCAACGTGAGATC 1731
Qy 432 TrpAspSerAsnMetLeuIleHisPheValLysAsnGlyProLysProLeuValTrpLeu 451
Db 1732 TGGCAAGATTGCGCGCTACTCATG-----AGCAAGCCCTTCATTGG--- 1776
Qy 452 PheValLysPheValSerLeuValLeuPheAsnLysAlaValLeuTrpPheGlyGly 471
Db 1777 -----GCGTTGTGAGCACCAAGATGTTGCGGTGGAAGCGCGCGGC 1818
Qy 472 ValProGlyLysGlnTyrAlaLeuIleLysAlaAsp----- 493
Db 1819 GTGCTCTGAACACCAAGCGGAATATGATGTTGAGCACTTCAAGAGACTGGGCTGACC 1878
Qy 484 -----GlyIleProIleGluAsnTyrIleAlaArg 493
Db 1879 AAGACATGACACCGGACAGGATATCAGTGGCGGTGCGATCTGCGAATATGTTGGAC 1938
Qy 494 ThrMetAspGlyValAlaGluAsnSerHisValArgLysGlnAsnTyrPheTyrAsn 513
Db 1939 ACCTGGACCCGTCGTGTTAACGAAACCATGATCAGCAACGATACTACTTTTACTTCTG 1998
Qy 514 CysLeuThrGlyLysPhe-----LeuArgAsp 522
Db 1999 TGTCTCAGGGGCGAGTTTTCAGACGAGTGCAAAACAAACCAATCAAGTGATCCGAGAT 2058
Qy 523 AsnCysProThrTyrLeuArgGluAlaPheAlaThrLeuLysSer---GlyValVal 541
Db 2059 GCCAC-CCACATACTTGTCTCCAGCCCATCTGAGCTCTTTGCCCGGTGCATTT 2117
Qy 542 AspAsnLeuThrValSerThrAsnPhePheMetGluLeuLys-----AlaArg 558
Db 2118 GACGGCTACGTATTACACAGATGAGATCAACGAGTGATCAAGCAATCACTCCCGCA 2177
Qy 559 ThrTyrThrLysValIleLeuMetAspHisValAspTrpLeuAspMetProValAlaAsn 578
Db 2178 AGCTGCAATTCGCTGTGATCATGACTCGATGACTGGTTGAT-----CCGGAAGGCACA 2234
Qy 579 GluLeu-----AlaGluCysLeuAlaLysGlnValAlaProGlyGlyValIle 595
Db 2235 GAAGCATGATCCAGGCCAGAACTCAACACGCCCTGAAATGGAGCGCGCATCTC 2294
Qy 596 TrpArgSerAlaSerLeuSerProProTyrAlaGluLeuIleGlnLysAlaGlyPheAsp 615
Db 2295 CTCCGCTCTGCAGCATCATGAGCCCTTGGTATATCAAGCATGTTTGAAGAGATGGTTTCA 2354
```

```
Qy 616 ValArgCysIle---ArgArgAlaThrGlnGlyTyrMetAspArgValAsnMetTyrSer 634
Db 2355 GCCCGCGGTGCGGCCCGCTCCCGGCTCTTGCATGATCGTGTGACATGATATGCA 2414
Qy 635 SerPheTyrMetAlaArgArg 641
Db 2415 TCGACGTGGATCTGTACCAAG 2435
```

## RESULT 6

```
ADW72751
ID ADW72751 standard; DNA; 3427 BP.
XX
AC ADW72751;
DT 21-APR-2005 (first entry)
XX
DE Neurospora crassa Btal gene.
XX
KW Phospholipid synthesis; transgenic plant; fertilizer; ds; gene;
KW betaine lipid.
XX
OS Neurospora crassa.
XX
PN PN PN PN PN PN PN PN PN PN PN PN PN PN PN PN PN PN PN PN PN
XX WO2005009115-A2.
XX
PD 03-FEB-2005.
XX
PF 15-JUL-2004; 2004WO-US022789.
XX
PR 16-JUL-2003; 2003US-00620914.
XX
PA (UNMS ) UNIV MICHIGAN STATE.
XX
PI Benning C, Riekhof W;
XX
DR WPI; 2005-112975/12.
DR P-PSDB; ADW72753.
XX
PT New composition comprising a purified DNA having an oligonucleotide
PT sequence encoding a protein, useful in producing Betaine lipid compounds
PT e.g., Diacylglycerol-O'-4'-(N,N,N-trimethyl) homoserine (DGTS).
XX
PS Example 5; SEQ ID NO 48; 147pp; English.
```

The invention relates to a new composition comprising a purified DNA having an oligonucleotide sequence appearing as ADW72747 (Chlamydomonas reinhardtii Btal cDNA) or ADW72752 (Neurospora crassa Btal coding region, encoding Btal proteins which are DGTS (diacylglycerol-O'-4'-(N,N,N-trimethyl)homoserine) synthetic enzymes. Btal (and BtaA and BtaB) enzymes allow the replacement of phospholipids with non-phosphorus containing lipids in the cell membrane. A transgenic plant expressing the enzymes would have a reduced need for phosphate-containing fertilizer. Also included are an RNA transcribed from the purified DNA, antibodies produced from the protein, a vector comprising the DNA or nucleic acid, a host cell comprising the vector, transgenic plants comprising the vector, and a protein translated from the RNA or encoded by the nucleic acid. The composition is useful in producing Betaine lipid compounds e.g., Diacylglycerol-O'-4'-(N,N,N-trimethyl) homoserine (DGTS). The composition of the invention reduces the amount of phosphate fertilizer needed for the optimal growth of crop plants. The present sequence encodes a betaine lipid synthetic enzyme.

XX SQ Sequence 3427 BP; 805 A; 938 C; 867 G; 814 T; 0 U; 3 Other;

## Alignment Scores:

Pred. No.:	5.92e-76	Length:	3427
Score:	804.00	Matches:	238
Percent Similarity:	43.2%	Conservative:	112
Best Local Similarity:	29.4%	Mismatches:	256
Query Match:	23.2%	Indels:	205
DB:	14	Gaps:	24



QY 568 isValaspTrpLeuaspMetPro-----Valalaa 578  
 Db : : : : :  
 2884 GTATGATTGGTTCGATCCGCTTCGCTGAGGAGGAAGGAGGAGGCGGCAAGCGGA 2943  
 QY 578 snGluLeuAlaGluCysLeuAlaLysGlnValalaproGlyGlylleValilleTrpArgS 598  
 Db : : : : :  
 2944 GGGAGCAAGTTCAGAGAGGTTGAATCGGGCGTTGAAGGTGGGTGGGAAGGTGTTGTTGAGGA 3003  
 QY 598 erAlaSerLeuSerProProTyrrAlaGluLeuileGlnLysAlaGlyPheAspValArgC 618  
 Db : : : : :  
 3004 GCGCGGAGTGGAGCGCGTGTATGTAGGGTTTTTGTGGAGGAAGGTTTGGAGCAAGGA 3063  
 QY 618 ysile-----ArgArgAlaThrGlnGlyTyrrMetAspArgValasnMet 633  
 Db : : : : :  
 3064 GGGTGGTTCGTGTAATCGGAAGGGGGACCAAGGAGTGTATTGACAGGTTGAATATGT 3123  
 QY 633 YrSerSerPheTyrrMetAlaArgArg 641  
 Db : : : : :  
 3124 ATGTAGTTGTTGATCTTGAAAAAG 3149

RESULT 7  
 ID ADR84882 standard; DNA; 2929 BP.  
 AC ADR84882;  
 DT 04-NOV-2004 (first entry)  
 XX Aspergillus fumigatus essential gene with introns #106.  
 DE Fungicide; Aspergillus fumigatus infection; Farmer's lung disease;  
 KW drug screening; ds.  
 XX Aspergillus fumigatus.  
 OS WO2004067709-A2.  
 PN 12-AUG-2004.  
 XX 16-JAN-2004; 2004WO-US001099.  
 XX 17-JAN-2003; 2003US-0441281P.  
 XX 13-JUN-2003; 2003US-0478196P.  
 XX (ELIT-) ELITRA PHARM INC.  
 XX (ELIT-) ELITRA CANADA LTD.  
 XX Jiang B, Hu W, Lemieux S, Roemer T;  
 XX WPI; 2004-594200/57.  
 DR P-PSDB; ADR86056.  
 XX New purified or isolated Aspergillus fumigatus nucleic acid molecule  
 PT encoding a gene product, useful for diagnosing and/or treating invasive  
 PT fungal infections, such as Farmer's lung disease.  
 XX Claim 3; SEQ ID NO 1106; 164pp; English.  
 XX The present invention relates to Aspergillus fumigatus genes that are  
 CC essential and are potential targets for drug screening. The methods and  
 CC compositions of the present invention are useful for diagnosing and/or  
 CC treating invasive Aspergillus fumigatus infection, including the allergic  
 CC forms of the disease, such as Farmer's lung disease. They can also be  
 CC used in various drug discovery purposes, such as expression of the  
 CC recombinant protein, hybridization assay and construction of nucleic acid  
 CC arrays. The present sequence represents an Aspergillus fumigatus  
 CC essential gene sequence containing all introns, used during diagnosis and  
 CC drug development in the invention. These genes share a high degree of  
 CC sequence conservation with known essential genes of candida albicans. The  
 CC sequence data for this patent is not represented in the printed  
 CC specification, but was obtained in electronic format from WIPO.  
 XX Sequence 2929 BP; 691 A; 791 C; 716 G; 731 T; 0 U; 0 Other;

Alignment Scores: 9.05e-65 Length: 2929  
 Pred. No.: 700.00 Matches: 219  
 Score: 38.0% Conservative: 94  
 Percent Similarity: 26.6% Mismatches: 232  
 Best Local Similarity: 20.2% Indels: 278  
 Query Match: 13 Gaps: 25  
 DB: 13

US-10-620-914-45 (1-648) x ADR84882 (1-2929)

QY 73 IleTrpValasp----- 76  
 Db 432 ATCTGGTTCGATTAAGACACCTCGGTATCGCTTATTGTTGCAAAAGCTGACACTCGTT 491  
 QY 77 -----LeuGlyGlyGlyThr-GlyGlu----- 83  
 Db 492 GCTTAGATTGGCGGAGGACCTGGTAAAGTGTCAATCGAGTCAACGCTTGTCTTCGAGGC 551  
 QY 84 -----AsnValaspMetMetAlaAspTyrrIleAspLeuAlaLysPh 97  
 Db 552 CTAAATCGTAGGATAGCTACACATTGAGGCCATGGCTTCTTCTTCCAGCTTACCAATT 611  
 QY 97 elysSer---IleTyrrValaspLeuCysHisSerLeuCysGluValalalalalalysLy 116  
 Db 612 CTTTCGACGCTGACCTTGTGACCTGTCTCCCTCTCTTTGCGAGGTTCCCGTCAGCG 671  
 QY 116 sAlaLysAlaLysGlyTyrrLysenValGlnValGluAlaAspAlaCysGlnPhe-- 135  
 Db 672 ATTCGAGCGCTAGGCTGGAAGAAATGTACGGTCTGTCAAGATGCGGCTCTTTCCG 731  
 QY 136 -----AlaProProGluGlyThrAlaThrLe 144  
 Db 732 TCTCCCCCATGAGAAAGTCGACCCCGTGCCTCCATCAACTGCAGCGCAGATCT 791  
 QY 144 uilleThrPheSerTyrrSerLeuThrMetIlePro----- 155  
 Db 792 TATCACCATGAGCTACAGCTTGTCCATGATTCTCGTGTAGTTTGTGATTAAACAAGCAT 851  
 QY 156 -----ProPheHisAsnValIleAspGlnAlaCysSerTy 167  
 Db 852 GAGGCGCGCTACTAAGTGAATACAGACTACTAGTGTCTGCTGACTTCCCTCCCTGACCGACCT 911  
 QY 167 rLeuSerGlnAspGlyLeuValGlyValAlaAspPhe----- 179  
 Db 912 CTTGAGGCGCTGCTGTTATCCTTGTCTGCGACTTTTATGTTAAGTAGCGTACAGCCT 971  
 QY 179 ----- 179  
 Db 972 CTGACGTGATAGATACTAATTTATCGTATGTCGCAAGTATCGTGGACGTTTCCGCCG 1031  
 QY 180 ----TyrrValSerGlyLysTyrrAspLeuProLeuArgGlnMetProTrpSerArgAspPh 198  
 Db 1032 CAACTACATTGGTGGTCTTTCAAT-----CGCACGTTAACTGGTGGTGGTGGTGG 1082  
 QY 198 ePheTrpArgSerIlePheAspIleAspAniIleAspIleGlyProGluArgArgAlaTy 218  
 Db 1083 CTTTGGCGGCGCTGGTTCGATCGGACCGTCTGCAAGCAGCAGCAGCAGCAGACTA 1142  
 QY 218 rLeuGluGlnLysLeuGluArgValTrpGluGlnAsnThrGln-----Glyse 234  
 Db 1143 TCTGGAATACCGCTTTGGAACCGCTCATCTCTGCGAGCGAGCGCAATTATCTTCTCGCGG 1202  
 QY 234 rIleProTyrrValProTrpLeuArgAlaProTyrrValTrpIleGlyArg----- 251  
 Db 1203 CATT-----CCTTACTACATCTTCTTGGACGTCAAGAA 1238  
 QY 251 ----- 251  
 Db 1239 TCTTGCATCAATCTCTCCGTCAGGAGACAATTGAATGTTGGACGCTCTCTTTACTGA 1298  
 QY 251 ----- 251



Db 1299 ATCTCCTTATCTTCTCCCGCAATCATCACCAAGGAGATGGAGAAAGCAATGGAGGAA 1358  
Qy 251 ----- 251  
Db 1359 CGCTCAAGAAATTCGTTCCAAAGCCTACGAGTGGCGGTTATATCAACCTGAGCGGAACCT 1418  
Qy 252 -----LeuProSerValGlyHisAlaLeuHisGluGluArgVal----- 264  
Db 1419 CCOCCTTCCCTCATCTTCTACCAAGATCAACCACTGCCGCATCTTCTACCAATGATCTGCT 1478  
Qy 265 -GluArgProMetPheProThrPheLeuTyrThrGlnSerTrpGluAspProG1 284  
Db 1479 GCCAAGCAGCAGCAGTTCAGATGATGATATATGCTTCACTTGGAGATCTCTCG 1538  
Qy 284 uProAspMetGluValMetGluLeuLeuAsnProLysAspThrValLeuThrLeuThrSerG1 304  
Db 1539 TGTGATCATCTCTCTCGACATCAAGCGGATGATGTTATCTTGGCCATTACCAAGCGC 1598  
Qy 304 yGlyCysAsnAlaLeuAsnLeuValGlnGlyAlaGlyGlnValValSerValAspCy 324  
Db 1599 CGGAGACATATTTTGGATTACCTTCCAGAGAGTCCACGAGAGTTCATGAGTCCGACCT 1658  
Qy 324 sAsnProAlaGlnSerAlaLeuLeuGluLeuLysValAlaLeuGlnLeuGluPh 344  
Db 1659 GAATCTTCAACAGATCATTTCTTGAATCTCAAGTTGCTAGTTTCATGGCCCTTGGTCA 1718  
Qy 344 eGluAspValTrpGlnLeuPheGlyGluGlyValHisProArgGlnGluLeuTyrG1 364  
Db 1719 TCGGATGTCTGGAAGATCTTTGGCGAGGAGAAACACCCAGAAATTCAGGGAATCTCTCAT 1778  
Qy 364 uLysLeuLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrpSerLysArgLeuTr 384  
Db 1779 TTCTCGTCTCAGCGCTCACCTTCCAGCCAGGATTCAGTACTGCTTGAGCACACTCA 1838  
Qy 384 pTyrPheGln-----HisGlyLeuTyrTrpGlnGlyMetGlyLysLeuCy 400  
Db 1839 CATTTTCACCTCAAAATATGGCAAGGACTTATGAACCGTGGCTCGCGCCACCGCAT 1898  
Qy 400 sTrpValLeuGlnCysLeuAlaValValLeuGlyLeuGlyLysThrValLysArgLeuAl 420  
Db 1899 CAAGATGGTTCCGTACCTGTTCAAGGTGTTGGCTTTGAGGTCAGGTGAAGAAGCTCTG 1958  
Qy 420 aAsnAlaProThrMetGluGlnArgArgLeuTrpAspSerAsnMetLeuLeuHisPh 440  
Db 1959 CGAGGCGCAGACTCTTCTGAGCAACAGTGAATCTGCGCAAGATTCGCGCGCTACTCAT 2018  
Qy 440 eValLysAsnGlyProLysProLeuValTrpLeuPheValLysPheValSerLeuValLe 460  
Db 2019 G-----AGCAAGCCCTTCATGG-----CGCGTTGT 2045  
Qy 460 uPheAsnLysAlaValLeuTrpPheGlyGlyGlyValProGlyLysGlnTyrAlaLeu11 480  
Db 2046 CAGCAGCAATGTTTCGCGTGAAGGCGCGGCGTGCCTCGAACCAGCGGATATGAT 2105  
Qy 480 eLysAlaAsp----- 483  
Db 2106 CGTTGACGACTACTTCAAGAGACTGGCGCTGACCAAGGACATGAACCCAGGCAAGGATAT 2165  
Qy 484 ----GlyLeuProLeuGlnAsnTyrLeuAlaArgThrMetAspGlyValAlaGluAsnSe 502  
Db 2166 CAGTGGCGGTCGATCTGGCAATATGTTGTGGACACACTGGACCCCGCTGTTAACGAAC 2225  
Qy 502 rHisValArgLysGlnAsnTyrPheTyrTrpAsnCysLeuThrGlyLysPheLeuArg-- 521  
Db 2226 CATGATCAGCAACGATAACTTACTTTTACTTCTCTGTTCTCCAGGCGCAGTTTCAAGACG 2285  
Qy 522 -----AspAsnCys---Pr 525  
Db 2286 GTGAGTTTGTCCCAAGAGTGCACAAACCAATCTAAGTGTATCCGAGATGCCACCC 2345  
Qy 525 oThrTyrLeuArgGluAlaAlaPheAlaThrLeuLysSer---GlyValValAspAsnLe 544  
Db 2346 CACATCTTCTCTCTCAGGCCCATGTGAAGCTCTCTTCTGCGCGGCTGATTTGACGGCT 2405

Qy 544 uThrValSerThrAsnPhePheMetGluGluLeuLys-----AlaArgThrTyrTh 561  
Db 2406 ACGTATTTACACAGATGAGATCAACGAAGTGATCAAGCAATCACTCCCGCAAGCCTGAC 2465  
Qy 561 rLysValLe----- 564  
Db 2466 AATCGTGGTGTAGCAGCGCTTGTATGATTTGCCAAATACCGCATCTAACCCATTG 2525  
Qy 565 ----LeuMetAspHisValAspTrpLeuAspMetProValAlaAsnGluLeu----- 580  
Db 2526 CCAGATCTAGGACTCCGATGGACTGGTTCGAT---CCGAAGGCACAGACATCGATCCA 2582  
Qy 581 -AlaGluCysLeuAlaLysGlnValAlaProGlyGlyLeuValLeuTrpArgSerAlaSe 600  
Db 2583 AGCCAGAAACTCAACCCAGCCCTGAAATGGACGCGCATTTCTCTCGCTCTGCGAG 2642  
Qy 600 rLeuSerProProTyrAlaGluLeuLeuGlnLysAlaGlyPheAspValArg----- 617  
Db 2643 CATCGAGCCTTGTGTATATCAAGCAGTTTCAAGAGATGGGTTTCACAGCCCGCTGTCGG 2702  
Qy 618 -----CysIleArg----- 621  
Db 2703 CGCCCGCTTCCCGGCTCTTGCATCGATCGTAAGTTCTCGATATTCCTCGACTGTTAC 2762  
Qy 622 ----Ala-ThrGlnGlyTyrMetAspArgValAsnMetTyrSerSerPheTyrMetAlaA 640  
Db 2763 CAATGCAATCGCTGACGCTAC---GACAGTGTGAACATGTATGATCATCGACGTGATCTGTA 2819  
Qy 640 rArg 641  
Db 2820 CCAAG 2824  
RESULT 8  
ID ADR84295 standard; DNA; 8929 BP.  
XX ADR84295;  
XX 04-NOV-2004 (first entry)  
DE Aspergillus fumigatus essential gene genomic sequence #106.  
KW Fungicide; Aspergillus fumigatus infection; Farmer's lung disease;  
KW drug screening; ds.  
XX Aspergillus fumigatus.  
XX WO2004067709-A2.  
XX 12-AUG-2004.  
XX 16-JAN-2004; 2004WO-US001099.  
PR 17-JAN-2003; 2003US-0441281P.  
PR 13-JUN-2003; 2003US-0478196P.  
XX (ELIT-) ELITRA PHARM INC.  
XX (ELIT-) ELITRA CANADA LTD.  
PI Jiang B, Hu W, Lemieux S, Roemer T;  
XX WPI; 2004-594200/57.  
DR P-PSDB; ADR86056.  
XX New purified or isolated Aspergillus fumigatus nucleic acid molecule  
PT encoding a gene product, useful for diagnosing and/or treating invasive  
PT fungal infections, such as Farmer's lung disease.  
XX Claim 3; SEQ ID NO 106; 164pp; English.  
XX The present invention relates to Aspergillus fumigatus genes that are  
CC essential and are potential targets for drug screening. The methods and



compositions of the present invention are useful for diagnosing and/or treating invasive *Aspergillus fumigatus* infection, including the allergic forms of the disease, such as Farmer's lung disease. They can also be used in various drug discovery purposes, such as expression of the recombinant protein, hybridization assay and construction of nucleic acid arrays. The present sequence represents an *Aspergillus fumigatus* essential gene full length genomic sequence, used during diagnosis and drug development in the invention. These genes share a high degree of sequence conservation with known essential genes of *Candida albicans*. The sequence data for this patent is not represented in the printed specification, but was obtained in electronic format from WIPO.

xx SQ Sequence 8929 BP; 2141 A; 2280 C; 2289 G; 2219 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 5.06e-64 Length: 8929  
Score: 700.00 Matches: 219  
Percent Similarity: 38.0% Conservative: 94  
Best Local Similarity: 26.6% Mismatches: 232  
Query Match: 20.2% Indels: 278  
DB: 13 Gaps: 25

US-10-620-914-45 (1-648) x ADR84295 (1-8929)

QY 73 IleTyrValAsp----- 76  
DB 3432 ATCTGGTGTATGTAAGACACCTCGGTATCGCTATTGTTGCGAAGCGCTGACACTCGTT 3491  
QY 77 -----LeuGlyGlyThr-GlyGlu----- 83  
DB 3492 GCTTAGATTGGCGGAGCAGCTGGTAAAGTGTCAAAATCGAGTCAACGCTTTGCTTCGAGGC 3551  
QY 84 -----AsnValAspMetMetAlaAspTyrIleAspLeuAlaLysPhe 97  
DB 3552 CTAATCGTAGATAGCTACAAACATTGAGGCGCATGCTTCTTCCACAGTTACCCCAATT 3611  
QY 97 eLysSer---IleTyrValValAspLeuCysHisSerLeuCysGluValAlaLysLysLy 116  
DB 3612 CTCTCGCAGGTGTACTTGTGACCTGTCTCCCTCTCTTTGCGAGGTTCGCGGTACAGG 3671  
QY 116 sAlaLysAlaLysGlyTrpLysAsnValGlnValValGluAlaAspAlaCysGlnPhe-- 135  
DB 3672 ATTCGAGCGCTAGGCTGGAAGATGTCAAGTGTCTGTCAAGATCGGCGCTTTTCGG 3731  
QY 136 -----AlaProGluGlyThrAlaThrLe 144  
DB 3732 TCTCCCCATGAGAAAGTCGACCCCGTGGCCAGCTCCATCAACTGCGAGCGCAGATCT 3791  
QY 144 uileThrPheSerTyrSerLeuThrMetIlePro----- 155  
DB 3792 TATCACCATGAGCTACAGCTTGTCATGATTTCTGTTAGTTTGGATTAAACAAGCATT 3851  
QY 156 -----ProPheHisAsnValIleAspGlnAlaCysSerTy 167  
DB 3852 GAGGCGCGCTACTAAGTGAATACAGACTACTACAGTGTCTGCTGACTCCCTCGACCGCCT 3911  
QY 167 rLeuSerGlnAspGlyLeuValGlyValAlaAspPhe----- 179  
DB 3912 CTTCGAGCGCTCTGGTATCTTGTGTCTGCGACTTTTATGGTAAGTAGCGTACAGCCTT 3971  
QY 179 ----- 179  
DB 3972 CTGACGTATAGATACTAATTTATCGTATAGTGCAAAGTATCTGGGACGTTTCGCGCCG 4031  
QY 180 ----TyrValSerGlyLysTyrAspLeuProLeuArgGlnMetProTyrSerArgPhe 198  
DB 4032 CAATACATTGGTGTGCTTTCAAT-----CGCCACGCTTAACCTGGCTGGGTCTGTC 4082  
QY 198 ePheTyrArgSerIlePheAspIleAspAsnIleAspIleGlyProGluArgAlaTy 218  
DB 4083 CTTTTGGCGGCCCTGGTTCGATGGCGCCGCTGTGAGTGTGGAAGCAGCAGCAGACTA 4142  
QY 218 rLeuGluGlnLysLeuGluArgValTrpGluGlnAsnThrGln-----Glyse 234

DB 4143 TCTGGAATACCGCTTTGGAAACCGTCTCTGCGAGGAGCGCAATATCTTCTCGCGCG 4202  
QY 234 rIleProTyrValProTyrIleuArgAlaProTyrTyrValTrpIleGlyArg----- 251  
DB 4203 CATT-----CCTTACTACATCTTCGTTGGAGCTCACAAGAA 4238  
QY 251 ----- 251  
DB 4239 TCTTGATCAATCTCTCGGTTCAGGAGACAATTGAATGTTGGACGCTTCTTTACTGA 4298  
QY 251 ----- 251  
DB 4299 ATCTCTTATCTTTCTCCGCAATCACACAGGAGAGTGGAGAAAGCAATGGAGAGAA 4358  
QY 251 ----- 251  
DB 4359 CGCTCAAGAAATTCGTTCCAAAGCCTACGAGTCGCGGTTATTCACCTGAGCGCGAACCT 4418  
QY 252 ----LeuProSerValGlyHisAlaLeuHisGluGluArgVal----- 264  
DB 4419 CCGCTTCCCTCATCTTCTACCAAGATCACCACTCGCGCATCTTCTACAATGATCTGCT 4478  
QY 265 -GluArgProProMetPheProThrPheLeuTyrThrGlnSerTyrGluAspProG1 284  
DB 4479 GCCAAGCACACGACGTTCAAGAAATGAATACATTTATGCTTCAACTGGGAAGATCCTCG 4538  
QY 284 uProAspMetGluValMetGluIleAsnProLysAspThrValLeuThrLeuThrSerG1 304  
DB 4539 TGTGATCATCTGCTTCTCGACATCAAGCGGATGATGTTATCTTGGCCATACACGCG 4598  
QY 304 yGlyCysAsnAlaLeuAsnLeuValGlnGlyAlaGlyGlnValValSerValAspCy 324  
DB 4599 CGAGACATATTTGGATTACCTTCAGAAGAGTCCACGAGAGTTTCATCGACTCGACCT 4658  
QY 324 sAsnProAlaGlnSerAlaLeuLeuGluLeuLysLysValAlaIleGlnLeuGluPh 344  
DB 4659 GAATCTCAACAGAAATCATTTGCTGAACCTCAAGGTGTGTAGTTTATGCGCCCTTGCTCA 4718  
QY 344 eGluAspValTrpGlnLeuPheGlyGluGlyValHisProArgIleGluGluLeuTyrG1 364  
DB 4719 TCAGATGTCTGAAGATCTTTGGCAGGGAACACCCAGAAATTCAGGGAATCTTCAT 4778  
QY 364 uLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrpSerLysArgLeuTr 384  
DB 4779 TTCTGCTCAGCGCTCACCTCTCCAGCAGGCAATCCAGTACTGCTGCGCCACACTCA 4838  
QY 384 pTyrPheGln-----HisGlyLeuTyrTyrGlnGlyGlyMetGlyLysLeuCy 400  
DB 4839 CATTTTCACTTCAAAATATGCAAAAGGACTTTATGAACCGGTGCTGCGCCACCGCCT 4898  
QY 400 sTrpValLeuGlnCysLeuAlaValValLeuGlyLeuGlyLysThrValLysArgLeuAl 420  
DB 4899 CAAGATGGTTCGGTACTCTGTTCAAGGTGTTGGCTTGAGGCTCAGGTGAAGAAGCTCTG 4958  
QY 420 aAsnAlaProThrMetGluGlnArgLeuTrpAspSerAsnMetLeuLeHisPh 440  
DB 4959 CGAGCGCAGACTCTTCTGAGCAACGTGAGATCTGGCCAAAGATTTCGCGCGTACTCAT 5018  
QY 440 eValLysAsnGlyProLysProLeuValTrpLeuPheValLysPheValSerLeuValLe 460  
DB 5019 G-----AGCAAGCCCTTCAATTGG-----CGGTGTGT 5045  
QY 460 uPheAsnLysAlaValLeuTrpPheGlyGlyValProGlyLysGlnTyrAlaLeuIl 480  
DB 5046 CAGCACCAATGGTTCGCTGGGAAGCGCGCGGCTGCTCGAAACACCGCGGAATATGAT 5105  
QY 480 eLysAlaAsp----- 483  
DB 5106 CGTTGACGACTACTTCAAGAGACTGGGCTGACCAAGGACATGAACCGAGCAAGGATAT 5165  
QY 484 ----GlyIleProIleGluAsnTyrIleAlaArgThrMetAspGlyValAlaGluAsnSe 502









Score: 329.00 Matches: 118  
Percent Similarity: 42.8% Conservative: 66  
Best Local Similarity: 27.4% Mismatches: 184  
Query Match: 9.5% Indels: 63  
DB: 14 Gaps: 15

US-10-620-914-45 (1-648) x ADW72725 (1-1248)

QY 248 TrpIleGlyArgLeuProSer-----ValGlyHisAlaLeuHisGluGluArgValGlu 265  
Db 48 TGAAGAGCGCGTACAGAAACCGCGCGCTTCCAAAGCGCGCATCTCCGAGCGCGTGT 107  
QY 266 ArgProMetPheProPro-ThrPheLeuTyrThrGlnSerTrpGluAspProGluPr 285  
Db 108 CGCTTCCTCTTTTCGGCGCTCGTC-----TATCCGAGATCTGGAGAACCCCGATGT 161  
QY 285 oAspMetGluValMetGluLeuAsnProLysAspThrValLeuThrLeuThrSerGlyG 305  
Db 162 CGACATGGAGCGCATCGAGCTTGTTCAGGGCCATCGCATCTGCATCTCTCCGCGG 221  
QY 305 YCysAsnAlaLeuAsnLeuValGlnGlyAlaGlyGlnValValSerValAspCysAs 325  
Db 222 CTGCAACATCTCGCTTACCTACCTACCGTTTCGGCGCGAGCATCGACGCGCTGCACTCA 281  
QY 325 nProAlaGlnSerAlaLeuLeuGluLeuLysValAlaIleGlnGlnLeuGluPheG 345  
Db 282 CGCGGCCACATCGCGTGAACCGCATGAAGCTGGAGCGCGTGCCTCTGCCCTCGCA 341  
QY 345 u---AspValTrpGlnLeuPheGlyGluGlyValHisProArgIleGluGluLeuTyrG 364  
Db 342 GGGCGATCTGTTTCGGCTTTTCGGCGCGCGACAGCCACCAATTCGACGCTATGA 401  
QY 364 uLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrpSerLysArgLeuTr 384  
Db 402 CGCTTTATTGCGCGCATCTCGATCGGTTCAGCGCGCACTATTGGAGCGCGCACTG 461  
QY 384 p-----TyrPheGlnHisGlyLeuTyrTyrGlnGlyMetG 397  
Db 462 GCGTGTGCGCGCGCATCGCGCTCTTCGACCGCAATTTCTACAGACCGCGCTGCTCGG 521  
QY 397 YLysLeuCysTrpValLeuGlnCysLeuAlaValValLeuGly----- 411  
Db 522 C-----CTGTTTCATCGCGCATGGCCATCGCACCGCGCAAAAT 557  
QY 412 -LeuGlyLysThrValLysArgLeuAlaAsnAlaProThrMetGluGluGlnArgLe 431  
Db 558 CTTTCGGCGTCAACCGCGCCACATGATGGAAGCCAGGAATATCGCGGAGCGCGCTT 617  
QY 431 uTrpAspSerAsnMetLeuIleHisPheValLysAsnGlyProLysProLeuValTrpLe 451  
Db 618 CTTCAACGAGGAGCTGGCGCGCTCTTCGACAAAG-----AAGCTTTTGAATGGGC 668  
QY 451 uPheValLysPheValSerLeuValLeuPheAsnLysAlaValLeuTrpPheGlyGly 471  
Db 669 GACCTCGGTAAGCGCTCGCTG-----TTCCGCTCGG 701  
QY 471 YValProGlyLysGlnTyr-----AlaLeuIleLysAlaAspGlyLeProIleGluAs 489  
Db 702 CATTTCGGCGCGCAGTACGATTCCTGATCCTCAGGCGAGCGCACCATCGCGCGT 761  
QY 489 nTyrIleAlaArg-----ThrMetAspGlyValAlaGluAsnSerHisValAr 505  
Db 762 TCTGAAGGCGCGCTGGAAGAGCTCGCTCGATTTTCCCTCGAAAC----- 810  
QY 505 gLysGlnAsnTyrPheTyrTrpAsnCysLeuThrGlyLysPheLeuArgAspAsn----- 523  
Db 811 -----AATTATTTCCTGGAGCGGCTTTTCGCCCGCTATTCAAATCCCGGTGAGGC 863  
QY 524 ----CysProThrTyrLeuArgGluAlaAlaPheAlaThrLeuLysSerGlyValValAs 542  
Db 864 CGCCCTCCCGCTATCTGGAAGAGCAGACATGACGAACCATCCGC-----GGCAATATCGA 920  
QY 542 pAsnLeuThrValSerThrAsnPhePheMetGluGluLeu-----LysAlaArgTh 559

Db 921 CCGCGTCCGCGCATCCACCATCCCAATCTGATCGAATTCCTCGCGGCAAGCAGCGGCAC 980  
QY 559 rTyrThrLysValIleLeuMetAspHisValAspTrpLeuAspMetProValAlaAsnG 579  
Db 981 CGTCGATCGCTTCATCTCTCGATCGGAGGACTGATGACCGATGACCATGCAACGC 1040  
QY 579 uLeuAlaGluCysLeuAlaLysGlnValAlaProGlyGlyLeValIleTrpArgSerAl 599  
Db 1041 GCTGTGTCGGAATCAGCGCACCGCTCCGCGAGCGCGCGCTCATCTTCGACCGC 1100  
QY 599 a-----SerLeuSerProProTyrAlaGluLeuIleGlnLysAlaGlyPheAsp-V 616  
Db 1101 CGCGAGCGCGCGCTCTCCACCGCGCTCTCGACCTCGCTCGACCGAGTGGGACTA 1160  
QY 616 aArgCysIleArgArgAlaThrGlnGlyTyrMetAspArgValAsnMetTyrSerSerP 636  
Db 1161 TCAGGACGAGCGCTCGCGGAATTCGCGCACGC-GACCTTCGGCCATCTATGGCGCT 1219  
QY 636 heTyrMetAlaArgArgLysGlyAla 644  
Db 1220 TCACCTCTATGTGAAGCGCACGGCA 1245

RESULT 13  
ABS58082  
ID ABS58082 standard; DNA; 1251 BP.  
XX  
AC ABS58082;  
XX  
DT 05-FEB-2003 (first entry)  
XX  
DE Agrobacterium tumefaciens *btA* gene.  
XX  
KW Betaine lipid production; betaine lipid compound; DGTS;  
KW diacylglycerol-0'-4'-(N,N,N'-trimethyl) homoserine; agricultural industry;  
KW phosphate-containing fertiliser; transgenic; *btA*; gene; ds.  
XX  
OS Agrobacterium tumefaciens.  
XX  
FH Location/Qualifiers  
FT 1..1251  
FT /tag= a  
FT /product= "BtA protein"  
XX  
XX W0200283844-A2.  
XX  
XX 24-OCT-2002.  
XX  
XX 09-APR-2002; 2002WO-US011134.  
XX  
XX 13-APR-2001; 2001US-0283812P.  
XX  
XX 08-APR-2002; 2002US-00283812.  
XX  
XX (UNMS ) UNIV MICHIGAN STATE.  
XX  
XX Benning C, Riekhof W, Klug R;  
XX WPI; 2003-058632/05.  
XX  
XX P-PSDB; ABG72178.  
XX  
XX New composition comprising an isolated and purified DNA molecule, useful  
XX for producing Betaine lipids, e.g. Diacylglycerol-0'-4'-(N,N,N'-trimethyl)  
XX homoserine (DGTS) for agricultural applications.  
XX  
XX Claim 1; Fig 24; 109pp; English.  
XX  
XX The present invention relates to compositions and methods for producing  
XX betaine lipids. The method comprises the expression of recombinant  
XX enzymes (e.g. from Rhodospirillum rubrum) in host cells such as  
XX bacteria, yeast, and plants to produce betaine lipid compounds including  
XX diacylglycerol-0'-4'-(N,N,N'-trimethyl) homoserine (DGTS). The methods and  
XX compositions of the invention are useful in agricultural applications, the  
XX such that the amount of phosphate-containing fertiliser required for the



CC growth of a particular plant is decreased. The polynucleotide sequences  
 CC encoding the recombinant enzymes may be used to produce vectors which can  
 CC be used to produce transgenic plants. The present sequence represents the  
 CC R. sphaeroides btaA gene homologue from Agrobacterium tumefaciens, btaA  
 XX  
 SQ Sequence 1251 BP; 252 A; 422 C; 345 G; 232 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 8, 51e-25 Length: 1251  
 Score: 326.50 Matches: 115  
 Percent Similarity: 47.6% Conservative: 71  
 Best Local Similarity: 29.4% Mismatches: 162  
 Query Match: 9.4% Indels: 46  
 DB: 8 Gaps: 14

US-10-620-914-45 (1-648) x ABS58082 (1-1251)

QY 275 LeuTyrThrGlnSerTrpGluAspProGluProAspMetGluValMetGluIleAsnPro 294  
 DB 130 GTCTATCCGAGATCTGGGAGACCCGAGATCGACATGGAAGCGATGGAGCTTGGCGAA 189  
 QY 295 LysAspThrValLeuThrLeuThrSerGlyGlyCysAsnAlaLeuAsnLeuValGln 314  
 DB 190 GGCCACCGCATCGTCACATCGCTCGCGCGCTGCAACATGCTGGCCTATCTCTCGCGC 249  
 QY 315 GlyAlaGlyGlnValSerValAspCysAsnProAlaGlnSerAlaLeuLeuGluLeu 334  
 DB 250 AACCGCGCAGCATCGATGCTGGAGCTTCAACCGCACATCGCGCTGAACAGCTG 309  
 QY 335 LysLysValAlaIleGlnGlnLeu---GluPheGluAspValTrpGlnLeuPheGlyGlu 353  
 DB 310 AAGCTCGCTCGCTCATCTGCGCGCCCATCAGGATGTGGTGGCCACTTTCGGCGCG 369  
 QY 354 GlyValHisProArgIleGlnGluLeu-TyrGluLysLysLeuAlaProPheLeuSerG1 373  
 DB 370 GCGCG-CACCGCGCAGCAACAGCGCTGATGACCGCTTTCATCGCGCAGCATCTGGATGC 428  
 QY 373 nThrSerHisAsnPheTrpSerLys-----ArgLeuTrpTyrPh 386  
 DB 429 CAGCAGCAAGCATCTACTGGTGAAGCGCACCCCTTTCGGCGCGCGTGCATTTTCGGTGT 488  
 QY 386 eGlnHisGlyLeuTyrTrpGlnGlyMetGlyLysLeuCysTrpValLeuGlnCysLe 406  
 DB 489 CGACAGGAACATCTACCGGACCGCGCTGCTCGCGGTTTCATCGCGCGCGCCACATCAT 548  
 QY 406 uAlaValValLeuGlyLeuGlyLysThrValLysArgLeuAlaAsnAlaProThrMetG1 426  
 DB 549 GGCCCGCTGCACGCGCTGAACTCACC-----GAAATGGCCCAAGACCGCGACGCTGGA 602  
 QY 426 uGluGlnArgArgLeuTrpAspSerAsnMetLeuIleHisPheValLysAsnGlyProLY 446  
 DB 603 CGAACAGCGCCAGTTTTCACACAGAGCGCGCGCTTTC-----GACAA 650  
 QY 446 sProLeuVal---TrpLeuPheValLysPheValSerLeuValLeuPheAsnLysAlaVa 465  
 DB 651 GCGGTGTGCTGGCTGGCTGACGAGCGCAAGAGCTCGCTT----- 690  
 QY 465 lLeuTrpPheGlyGlyGlyValProGlyLysGlnTyr-----AlaLeuIleLysAl 482  
 DB 691 -----TTGCGCTTGGCATTTCCGCGCGCGCATATGACGAGCTTTCAG 743  
 QY 482 aAspGlyLeuProIleGluAsnTyrIleAlaArgThrMetAspGlyValAlaGluAsnSe 502  
 DB 744 CGACGGC---ACGTTGCTCCCTCCCTCAAGGACCGCTGGAAAGCTTGCCTCACTT 800  
 QY 502 rHisValArgLysGlnAsnTyrPheTyrAsnCysLeuThrGlyLysPheLeuArgAs 522  
 DB 801 CCCGCTC---AGCGACAATATTTCGCTGGCAGCGCTTTTCGGCGCGCTTATCCGAGCC 857  
 QY 522 pAsn-----CysProThrTyrLeuArgGluAlaAlaPheAlaThrLeuLysSerG1 539  
 DB 858 GCATGAGGGTGCCTGCCCTTATCTCAAGCGCGGAATATTACGAAGAAGATCCGCAACA 917

QY 539 yValValAspAsnLeuThrValSerThrAsnPhePheMetGluGluLeuLys----- 556  
 DB 918 CACCGCG---CGGTGCGGTGATCATCGCCACCTATACGAGCTGCTTTCCCGCAAGCC 974  
 QY 557 -AlaArgThrTyrThrLysValIleLeuMetAspHisValAspTrpLeuAspMetProVa 576  
 DB 975 GGCNAATGGCGTGCAGCCGTATATCTCTGCTGATGCGGAGGACTGGATGATGCGATGTGCA 1034  
 QY 576 lAlaAsnGluLeuAlaGluCysLeuAlaLysGlnValAlaProGlyGlyIleValIleTr 596  
 DB 1035 GCTCAACGAGTATATGTCGAGATCAGCCGACTGCGGATCTCCGGGCGAGCGTCACTTT 1094  
 QY 596 pArgSerAlaSerLeu-----SerProProTyrAlaGluLeuIleGlnLY 611  
 DB 1095 CCGCACCGCGCGCA-AAAGAGCGTTATCGAGGCGCGCTTTCGCGGACATCCGCAACC 1153  
 QY 611 sAlaGlyPheAspValArgCysIleArgArgAlaThrGlnGlyTyrMetAspArgValAs 631  
 DB 1154 AGTGGGTCTATCTCGAAGAGCGCTCCAACGA---ACTCAAC--GCCATGGACCGCTCGGC 1208  
 QY 631 nMetTyrSerSerPheTyrMetAlaArgArg 641  
 DB 1209 CATTTATGGCGGTTCCTCATATCTACCAGAGG 1239  
 RESULT 14  
 ADW72731  
 ID ADW72731 standard; DNA; 1251 BP.  
 XX AC ADW72731;  
 XX 21-APR-2005 (first entry)  
 DT XX Agrobacterium tumefaciens btaA gene.  
 DE XX  
 XX Phospholipid synthesis; transgenic plant; fertilizer; ds; gene;  
 KW betaine lipid.  
 KW Agrobacterium tumefaciens.  
 OS  
 XX Key Location/Qualifiers  
 FH CDS 1.1251  
 FT /\*tag= a  
 FT /product= "BtaA protein"  
 XX  
 XX WO2005009115-A2.  
 XX 03-FEB-2005.  
 XX 15-JUL-2004; 2004WO-US022789.  
 XX 16-JUL-2003; 2003US-00620914.  
 XX (UNMS ) UNIV MICHIGAN STATE.  
 XX Benning C, Riekhof W;  
 XX WPI: 2005-112975/12.  
 XX P-PSDB; ADW72732.  
 XX  
 XX New composition comprising a purified DNA having an oligonucleotide  
 XX sequence encoding a protein, useful in producing Betaine lipid compounds  
 XX e.g., Diacylglycerol-O-4'-(N,N,N-trimethyl) homoserine (DGTS).  
 XX Disclosure; SEQ ID NO 28; 147pp; English.  
 XX  
 XX The invention relates to a new composition comprising a purified DNA  
 XX having an oligonucleotide sequence appearing as ADW72747 (Chlamydomonas  
 XX reinhardtii BtaI cDNA) or ADW72752 (Neurospora crassa BtaI coding region,  
 XX encoding BtaI proteins which are DGTS (diacylglycerol-O-4'-(N,N,N-  
 XX trimethyl)homoserine) synthetic enzymes. BtaI (and BtaA and BtaB) enzymes  
 XX allow the replacement of phospholipids with non-phosphorus containing  
 XX lipids in the cell membrane. A transgenic plant expressing the enzymes  
 XX would have a reduced need for phosphate-containing fertilizer. Also



CC included are an RNA transcribed from the purified DNA, antibodies  
 CC produced from the protein, a vector comprising the DNA or nucleic acid, a  
 CC host cell comprising the vector, transgenic plants comprising the vector,  
 CC and a protein translated from the RNA or encoded by the nucleic acid. The  
 CC composition is useful in producing Betaine lipid compounds e.g.,  
 CC Diacylglycerol-0'-4'-(N,N,N-trimethyl) homoserine (DGTS). The composition  
 CC of the invention reduces the amount of phosphate fertilizer needed for  
 CC the optimal growth of crop plants. The present sequence encodes a betaine  
 CC lipid synthetic enzyme.

XX  
 SQ Sequence 1251 BP; 252 A; 423 C; 345 G; 231 T; 0 U; 0 Other;

#### Alignment Scores:

Pred. No.: 8,51e-25 Length: 1251  
 Score: 326.50 Matches: 115  
 Percent Similarity: 47.6% Conservative: 71  
 Best Local Similarity: 29.4% Mismatches: 162  
 Query Match: 9.4% Indels: 46  
 DB: 14 Gaps: 14

US-10-620-914-45 (1-648) x ADW72731 (1-1251)

QY 275 LeuTyrThrGlnSerTrpGluAspProGluProAspMetGluValMetGluIleAsnPro 294  
 DB 130 GTCTATCCGAGATCTGGAGAGACCCGAGATCGACATGAGGATGGAGCTTGGCGAA 189  
 QY 295 LysAspThrValLeuThrLeuThrSerGlyGlyCysAsnAlaLeuAsnLeuValGln 314  
 DB 190 GGCACCGCATCTCACCATCGCTCGCGGCTGCAACATGCTGGCTATCTCTCGCGC 249  
 QY 315 GlyAlaGlyGlnValSerValAspCysAsnProAlaGlnSerAlaLeuLeuGluLeu 334  
 DB 250 AACCCGCGCAGCATCGATGGTGGAGCCTCAACCCGACACATCGCGCTGAACAAGCTG 309  
 QY 335 LysLysValAlaIleGlnGlnLeu---GluPheGluAspValTrpGlnLeuPheGlyGlu 353  
 DB 310 AAGCTCGCTCGCTTCGCCATCTGCCGCCATCAGATGTGTGGCCATTCGCGCGC 369  
 QY 354 GlyValHisProArgIleGluGluLeu-TyrGluLysLysLeuAlaProPheLeuSerG1 373  
 DB 370 GCCGG-CACCCGCGAGCAACAGCGTCGTTATGACCGTTTCATCGCGAGCATCTGGATGC 428  
 QY 373 nThrSerHisAsnPheTrpSerLys-----ArgLeuTrpTyrPh 386  
 DB 429 CAGCAACAAGCATACTGTGTCGAAGCGCACCTTTCCGCGCGCTGCGCATTTCCGTGT 488  
 QY 386 eGlnHisGlyLeuTyrTrpGlnGlyMetGlyLysLeuCysTrpValLeuGlnCysLe 406  
 DB 489 CGACAGACATCTACCGGACCGGCTGCTCGCGCGTTTCATCGCGCGCGCCACATCAT 548  
 QY 406 uAlaValLeuGlyLeuGlyLysThrValLysArgLeuAlaAsnAlaProThrMetG1 426  
 DB 549 GGGCCGCTGCACGGCGTGAATCACC-----GAAATGGCCAAAGACCGGAGCGTGA 602  
 QY 426 uGluGlnArgArgLeuTrpAspSerAsnMetLeuIleHisPheValLysAsnGlyProly 446  
 DB 603 CGAACAGCGCCAGCTTTTTCGACAGAGGTGCGCGCTTTTC-----GACAA 650  
 QY 446 sProLeuVal---TrpLeuPheValLysPheValSerLeuValLeuPheAsnLysAlaVa 465  
 DB 651 GCGGTGTGTGCTGCTGCTGACGAGCGCAAGAGCTCGCTT----- 690  
 QY 465 lLeuTrpPheGlyGlyValProGlyLysGlnTyr-----AlaLeuIleLysAl 482  
 DB 691 -----TTCCGGCTTGGCATTCGCGCGCGCAGTATGACGAGCTGGCAAGCTTTCCAG 743  
 QY 482 aAspGlyIleProIleGluAsnTyrIleAlaArgThrMetAspGlyValAlaGluAsnSe 502  
 DB 744 CGACGGC---ACGTTTGCCTCCCTCAAGAGCGGCTGGAAAGCTTGGCTGCAACTT 800  
 QY 502 rHisValArgLysGluAsnTyrPheTyrTrpAsnCysLeuThrGlyLysPheLeuArgAs 522  
 DB 801 CCGGCTC---ACGACAAATATTTCCTCGCTGGCAGGCGCTTTGGCGCGCTTATCCGAGCC 857

QY 522 pAsn-----CysProThrTyrLeuArgGluAlaAalaPheAlaThrLeuLysSerG1 539  
 DB 858 GCATAGGGTGGCTCGCGCTTATCTCAAGCCGGAATATTACGAAAGATCCGCAACA 917  
 QY 539 yValValAspAsnLeuThrValSerThrAsnPhePheMetGluGluLeuLys----- 556  
 DB 918 CACCGCG---CGCGTGGCGGTGCATCAGCCACCTATACCGAGCTGCTTTCCCGCAAGCC 974  
 QY 557 -AlaArgThrTyrThrLysValIleLeuMetAspHisValAspTrpLeuAspMetProVa 576  
 DB 975 GGCAATGGCTCGACCGCTATATCTGCTCGATGGCAGACTGGATGACGATGTGCA 1034  
 QY 576 lAlaAsnGluLeuAlaGluCysLeuAlaLysGlnValAlaProGlyGlyIleValIleTr 596  
 DB 1035 GCTCAACGAGTATGTGTCGAGATCAGCGCACTGCGGCATCGGGGACGCGTCATCTT 1094  
 QY 596 pArgSerAlaSerLeu-----SerProTyrAlaGluLeuLeuGlnLys 611  
 DB 1095 CCGCACCGCGCGCGA-AAAGAGCGTTATCGAGGCGCGCTTTCGCGCGACATCCGCAACC 1153  
 QY 611 sAlaGlyPheAspValArgCysIleArgArgAlaThrGlnGlyTyrMetAspArgValAs 631  
 DB 1154 AGTGGTCTATCTCGAAGAGCGCTCCACGA---ACTCAAC---GCCATGACCGCTCGGC 1208  
 QY 631 nMetTyrSerSerPheTyrMetAlaArgArg 641  
 DB 1209 CATTTATGGCGGCTTCCATATCTACCAGAGG 1239  
 RESULT 15  
 ABS58084  
 ID ABS58084 standard; DNA; 1251 BP.  
 XX  
 AC ABS58084;  
 XX  
 DT 05-FEB-2003 (first entry)  
 XX  
 DE Sinorhizobium meliloti btaA gene.  
 XX  
 KW Betaine lipid production; betaine lipid compound; DGTS;  
 KW diacylglycerol-0'-4'-(N,N,N-trimethyl) homoserine; agricultural industry;  
 KW phosphate-containing fertilizer; transgenic; btaA; gene; ds.  
 XX  
 OS Sinorhizobium meliloti.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..1251  
 FT /\*tag= a  
 FT /product= "Bta protein"  
 XX  
 PN WO200283844-A2.  
 XX  
 PD 24-OCT-2002.  
 XX  
 PF 09-APR-2002; 2002WO-US011134.  
 XX  
 PR 13-APR-2001; 2001US-0283812P.  
 PR 08-APR-2002; 2002US-00283812.  
 XX  
 PA (UNWS ) UNIV MICHIGAN STATE.  
 XX  
 PI Benning C, Riekhof W, Klug R;  
 XX  
 DR WPI; 2003-058632/05.  
 DR P-PSDB; ABG72180.  
 XX  
 PT New composition comprising an isolated and purified DNA molecule, useful  
 PT for producing Betaine lipids, e.g. Diacylglycerol-0'-4'-(N,N,N-trimethyl)  
 PT homoserine (DGTS) for agricultural applications.  
 XX  
 PS Claim 1; Fig 28; 109pp; English.  
 XX  
 CC The present invention relates to compositions and methods for producing

CC betaine lipids. The method comprises the expression of recombinant  
CC enzymes (e.g. from *Rhodospirillum rubrum*) in host cells such as  
CC bacteria, yeast, and plants to produce betaine lipid compounds including  
CC diacylglycerol-0'-*(N,N,N*-trimethyl) homoserine (DGTS). The methods and  
CC compositions of the present invention are useful in agricultural applications,  
CC such that the amount of phosphate-containing fertiliser required for the  
CC growth of a particular plant is decreased. The polynucleotide sequences  
CC encoding the recombinant enzymes may be used to produce vectors which can  
CC be used to produce transgenic plants. The present sequence represents the  
CC R. *sphaeroides* braA gene homologue from *Sinorhizobium meliloti*, braA  
XX Sequence 1251 BP; 227 A; 440 C; 355 G; 229 T; 0 U; 0 Other.

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 13, 2006, 19:55:21 ; Search time 1196 Seconds  
(without alignments)  
10849.629 Million cell updates/sec

Title: US-10-620-914-44

Perfect score: 1947

Sequence: 1 atggggctgggtcgtgacgg.....gcccgaagaagacaactaa 1947

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_21.\*

1: Geneseqn1980s.\*

2: Geneseqn1990s.\*

3: Geneseqn2000s.\*

4: Geneseqn2001as.\*

5: Geneseqn2001bs.\*

6: Geneseqn2002as.\*

7: Geneseqn2002bs.\*

8: Geneseqn2003as.\*

9: Geneseqn2003bs.\*

10: Geneseqn2003cs.\*

11: Geneseqn2003ds.\*

12: Geneseqn2004as.\*

13: Geneseqn2004bs.\*

14: Geneseqn2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1947	100.0	1947	14	Adw72747
2	645	33.1	5242	14	Adw72746 Chlamydom
3	100.2	5.1	1252	8	Abss58051 Gene enco
4	100.2	5.1	1252	14	Adw72704 Rhodobact
5	76.8	3.9	2259	12	Adp98536 C. albica
6	73	3.7	2727	14	Adw72752 Neurospor
7	73	3.7	3427	14	Adw72751 Parapoxvi
8	72.4	3.7	137560	12	Abss58084 Sinorhizo
9	70	3.6	1251	8	Abss58076 Mesorhizo
10	69.6	3.6	1248	14	Adw72735
11	69.6	3.6	1248	14	Adw72725 Mesorhizo
12	69.4	3.6	1092	4	Aah44047 Streptomy
13	69.4	3.6	3003	11	Abd16186 Pseudomon
14	69.4	3.6	3132	11	Abd15936 Pseudomon
15	69.4	3.6	3411	11	Abd16219 Pseudomon
16	69.4	3.6	1092	6	AbL41287 Streptomy
17	67.8	3.5	1104	5	Aah78257 Nucleotid
18	67.8	3.5	1104	6	AbL41308 Streptomy
19	67.8	3.5	1104	6	AbL41308 Streptomy

c	20	67.8	3.5	2000	11	ACL36843	ACL36843 Rice stre
	21	67.8	3.5	6798	4	Aah44043	Aah44043 Streptomy
	22	67.8	3.5	6798	5	Aah78258	Aah78258 Nucleotid
	23	67.8	3.5	6798	6	Ad31022	Ad31022 Streptomy
	24	67.8	3.5	6798	12	Adp90610	Adp90610 Streptomy
	25	67.8	3.5	6798	13	Adr16784	Adr16784 Streptomy
	26	67.8	3.5	8077	6	Ad31028	Ad31028 Operon F
	27	67.8	3.5	8719	13	Adp90614	Adp90614 Streptomy
	28	67.8	3.5	8719	13	Adr16788	Adr16788 Streptomy
	29	67.6	3.5	5858	8	Ad55726	Ad55726 Nephila m
	30	66.6	3.4	1626	12	Ad051196	Ad051196 Staphyloc
	31	66.6	3.4	1626	12	Ad050828	Ad050828 S. coelic
	32	66.6	3.4	1626	13	Ad84636	Ad84636 Streptomy
	33	66.6	3.4	2535	14	Ad204182	Ad204182 Modified
	34	65	3.3	11238	10	Ad55817	Ad55817 Micromono
	35	65	3.3	47988	10	Ad86070	Ad86070 Streptomy
	36	65	3.3	60196	10	Ad55810	Ad55810 Micromono
	37	64.8	3.3	2583	14	Ad207737	Ad207737 HIV subty
	38	64.6	3.3	2034	14	AdY28052	AdY28052 PERV subc
	39	64.2	3.3	2278	14	AdY57110	AdY57110 Zea mays
	40	64.2	3.3	2807	13	AdX62961	AdX62961 Plant ful
	41	64	3.3	1005	2	AAZ30006	AAZ30006 Optimised
	42	64	3.3	1076	2	AAZ30007	AAZ30007 Optimised
	43	63	3.2	1891	14	ACL70101	ACL70101 M. xanthu
	44	63	3.2	2034	14	AdY28053	AdY28053 PERV subc
	45	63	3.2	2034	14	ACL72051	ACL72051 M. xanthu

ALIGNMENTS

RESULT 1

ADW72747

ID ADW72747 standard; cdna; 1947 bp.

XX ADW72747;

AC ADW72747;

XX 21-APR-2005 (first entry)

XX Chlamydomonas reinhardtii Btal cdna.

DE Chlamydomonas reinhardtii

XX Phospholipid synthesis; transgenic plant; fertilizer; ss; gene;

XX betaine lipid.

XX Chlamydomonas reinhardtii.

OS Chlamydomonas reinhardtii.

XX Key Location/Qualifiers

FH 1..1947

FT /\*tag= a

FT /product= "Btal protein"

XX WO2005009115-A2.

XX 03-FEB-2005.

XX 15-JUL-2004; 2004WO-US022789.

XX 16-JUL-2003; 2003US-00620914.

XX (UNMS ) UNIV MICHIGAN STATE.

XX Benning C, Riekhof W;

XX WPI: 2005-112975/12.

XX P-PSDB; ADW72748.

XX New composition comprising a purified DNA having an oligonucleotide

PT sequence encoding a protein, useful in producing Betaine lipid compounds

PT e.g., Diacylglycerol-O-4'-(N,N,N-trimethyl) homoserine (DGTS).

XX Claim 1; SEQ ID NO 44; 147pp; English.

XX The invention relates to a new composition comprising a purified DNA

CC

CC having an oligonucleotide sequence appearing as ADW72747 (Chlamydomonas  
CC reinhardtii Btal cDNA) or ADW72752 (Neurospora crassa Btal coding region,  
CC encoding Btal proteins which are DGTS (diacylglycerol-O-4'-(N,N,N,-  
CC trimethyl)homoserine) synthetic enzymes. Btal (and BtaA and BtaB) enzymes  
CC allow the replacement of phospholipids with non-phosphorus containing  
CC lipids in the cell membrane. A transgenic plant expressing the enzymes  
CC would have a reduced need for phosphate-containing fertilizer. Also  
CC included are an RNA transcribed from the purified DNA, antibodies  
CC produced from the protein, a vector comprising the DNA or nucleic acid, a  
CC host cell comprising the vector, transgenic plants comprising the vector,  
CC and a protein translated from the RNA or encoded by the nucleic acid. The  
CC composition is useful in producing Betaine lipid compounds e.g.,  
CC Diacylglycerol-O-4'-(N,N,N,-trimethyl) homoserine (DGTS). The composition  
CC of the invention reduces the amount of phosphate fertilizer needed for  
CC the optimal growth of crop plants. The present sequence encodes a betaine  
CC lipid synthetic enzyme.

XX  
SQ Sequence 1947 BP; 365 A; 619 C; 613 G; 350 T; 0 U; 0 Other;

Query Match 100.0%; Score 1947; DB 14; Length 1947;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1947; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGGGGTTCGGGTCTGACGGCCGGCCCTGCGAGCTACACCAAGAGAACTTCTCCCTGGAG 60  
DB 1 ATGGGGTTCGGGTCTGAGGCGCGCCCTGCGAGCTACACCAAGAGAACTTCTCCCTGGAG 60  
QY 61 AAGCTCAAGCTCAGCAGATGAAGGATGACCTGACCTGACCTGACCTGACCTGACCTGACCT 120  
DB 61 AAGCTCAAGCTCAGCAGATGAAGGATGACCTGACCTGACCTGACCTGACCTGACCTGACCT 120  
QY 121 AGCAAGAGGCGATGATACGCTGCTGCGCTGCGAGCTTCTACGGGCGCCAGGCGCT 180  
DB 121 AGCAAGAGGCGATGATACGCTGCTGCGCTGCGAGCTTCTACGGGCGCCAGGCGCT 180  
QY 181 GCCTTTGCTCCGCGCTGCGAGCGCTCGAACTCATCTGGGTGACCTGGGTGGTGGC 240  
DB 181 GCCTTTGCTCCGCGCTGCGAGCGCTCGAACTCATCTGGGTGACCTGGGTGGTGGC 240  
QY 241 ACTGGGAGAAATGTCATATGATGGCTGATTAATCATGACCTGGCGAAAGTTCAAGTCCATC 300  
DB 241 ACTGGGAGAAATGTCATATGATGGCTGATTAATCATGACCTGGCGAAAGTTCAAGTCCATC 300  
QY 301 TAGTGTGTCACCTGTCGCTGCTGCGAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360  
DB 301 TAGTGTGTCACCTGTCGCTGCTGCGAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360  
QY 361 GGCTGGAAGAAATGTCAGGTGCTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420  
DB 361 GGCTGGAAGAAATGTCAGGTGCTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420  
QY 421 ACCGCGACGCTCATCACTTCTCTA CTGCGTCAAGATGATTCACCGTTCCACAAAGTCC 480  
DB 421 ACCGCGACGCTCATCACTTCTCTA CTGCGTCAAGATGATTCACCGTTCCACAAAGTCC 480  
QY 481 ATCGACACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540  
DB 481 ATCGACACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540  
QY 541 GTGACGGGCAAGTACGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600  
DB 541 GTGACGGGCAAGTACGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600  
QY 601 CGATCGATCTTCGACATCGACAAATTTGACATCGGCGCGGCGGCGGCGGCGGCGGCGGCGG 660  
DB 601 CGATCGATCTTCGACATCGACAAATTTGACATCGGCGCGGCGGCGGCGGCGGCGGCGGCGG 660  
QY 661 CAGAGCTGGAGCGCGTGTGGGAGCAGACACACAGAGTTTCAATCCCTAGTGGCGCTGG 720  
DB 661 CAGAGCTGGAGCGCGTGTGGGAGCAGACACACAGAGTTTCAATCCCTAGTGGCGCTGG 720  
QY 721 CTGCGCGCCCTCACTACGCTGTGATTTGGCGCGCTGCGCGGCTTGGCGGCGGCGGCGGCGG 780  
DB 721 CTGCGCGCCCTCACTACGCTGTGATTTGGCGCGCTTGGCGGCGGCGGCGGCGGCGGCGGCGG

DB 721 CTGCGCGCCCTCACTACGCTGTGATTTGGCGCGCTGCGCGGCGGCGGCGGCGGCGGCGGCGG 780  
QY 781 GAGGAGCGGTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840  
DB 781 GAGGAGCGGTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840  
QY 841 GAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 900  
DB 841 GAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 900  
QY 901 CTGACTAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960  
DB 901 CTGACTAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960  
QY 961 TCGGTGAGCTGCAACCCCGCGAGTGGCGGCTTCTGAGCTGAAGAGGTGGCGGCGGCGGCGG 1020  
DB 961 TCGGTGAGCTGCAACCCCGCGAGTGGCGGCTTCTGAGCTGAAGAGGTGGCGGCGGCGGCGG 1020  
QY 1021 CAGCTGAGGTTTTCAGGACGCTGTGGCGAGCTGTTTCGGGAGGCGGCGGCGGCGGCGGCGGCGG 1080  
DB 1021 CAGCTGAGGTTTTCAGGACGCTGTGGCGAGCTGTTTCGGGAGGCGGCGGCGGCGGCGGCGG 1080  
QY 1081 GAGCTGTACGAGAAAGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1140  
DB 1081 GAGCTGTACGAGAAAGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1140  
QY 1141 AAGGCGCTCTGGTACTTCCAGCACGCGCTGTACTACAGGCGGCGGCGGCGGCGGCGGCGGCGG 1200  
DB 1141 AAGGCGCTCTGGTACTTCCAGCACGCGCTGTACTACAGGCGGCGGCGGCGGCGGCGGCGGCGG 1200  
QY 1201 TGGGTGCTGCAAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1260  
DB 1201 TGGGTGCTGCAAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1260  
QY 1261 AACGCGCCCAAAATGAGGAGAGCAGCGCGTCTGTGGGAGCAGCAACATGCTCATCATTTC 1320  
DB 1261 AACGCGCCCAAAATGAGGAGAGCAGCGCGTCTGTGGGAGCAGCAACATGCTCATCATTTC 1320  
QY 1321 GTGAAGAACGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1380  
DB 1321 GTGAAGAACGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1380  
QY 1381 TTCAAAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1440  
DB 1381 TTCAAAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1440  
QY 1441 AAGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1500  
DB 1441 AAGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1500  
QY 1501 AACTCGCACGTCGCGAGCAGAACTACTTCTACTACAACTGCTCTACCGGCGGCGGCGGCGGCGG 1560  
DB 1501 AACTCGCACGTCGCGAGCAGAACTACTTCTACTACAACTGCTCTACCGGCGGCGGCGGCGGCGG 1560  
QY 1561 CGCGACAACTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1620  
DB 1561 CGCGACAACTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1620  
QY 1621 GTGGAACAACTGACCGGCTTCCACAACTTCTTATGAGGAGGCTCAAGGCGGCGGCGGCGGCGG 1680  
DB 1621 GTGGAACAACTGACCGGCTTCCACAACTTCTTATGAGGAGGCTCAAGGCGGCGGCGGCGGCGG 1680  
QY 1681 ACCAAGGTGATTCGATGGACCGCTGGAATGCGGCTGATATGCCCGTGGCGGCGGCGGCGGCGG 1740  
DB 1681 ACCAAGGTGATTCGATGGACCGCTGGAATGCGGCTGATATGCCCGTGGCGGCGGCGGCGGCGG 1740  
QY 1741 GCGGAGTGCCTGGCGCAAGCAGGTTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1800  
DB 1741 GCGGAGTGCCTGGCGCAAGCAGGTTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1800  
QY 1801 CTAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1860  
DB 1801 CTAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1860

	1920	1947	1947
Qy	CGGCCACTCAGGGCTACATGACCGCGTCAA	CGGCCACTCAGGGCTACATGACCGCGTCAA	CGGCCACTCAGGGCTACATGACCGCGTCAA
Db	CGGCCACTCAGGGCTACATGACCGCGTCAA	CGGCCACTCAGGGCTACATGACCGCGTCAA	CGGCCACTCAGGGCTACATGACCGCGTCAA
Qy	CGGCCACTCAGGGCTACATGACCGCGTCAA	CGGCCACTCAGGGCTACATGACCGCGTCAA	CGGCCACTCAGGGCTACATGACCGCGTCAA
Db	CGGCCACTCAGGGCTACATGACCGCGTCAA	CGGCCACTCAGGGCTACATGACCGCGTCAA	CGGCCACTCAGGGCTACATGACCGCGTCAA

## RESULT 2

ADW72746  
ID ADW72746 standard; DNA; 5242 BP.

ADW72746:

DT 21-APR-2005 (first entry)

XX  
DE Chlamydomonas reinhardtii Btal gene.

XX  
KW Phospholipid synthesis; transgenic plant; fertilizer; ds; gene;  
KW betaine lipid.

XX Chlamydomonas reinhardtii.

XX  
PN WO2005009115-A2

XX  
PD 03-FEB-2005.

XX  
PF 15-JUL-2004; 2004WO-US022789.

XX  
PR 16-JUL-2003; 2003US-00620914.

XX  
PA (UNMS ) UNIV MICHIGAN STATE.

XX  
PI Benning C, Riekhof W;

XX  
DR WPI; 2005-112975/12.

DR P-PSDB; ADW72748.  
XX

New composition comprising a purified DNA having an oligonucleotide sequence encoding a protein, useful in producing Betaine lipid compounds e.g., Diacylglycerol-0-4'-(N,N,N'-trimethyl) homoserine (DGTS).

XX  
PS Example 4; SEQ ID NO 43; 147pp; English.

XX The invention relates to a new composition comprising a purified DNA  
CC having an oligonucleotide sequence appearing as ADW27477 (*Chlamydomonas*  
CC reinhardtii Bsal cDNA) or ADW2752 (*Neurospora crassa* Bsal cloning region,  
CC encoding Bsal proteins) which are DGTS (diacylglycerol-O-4'-(N,N,N',N'-  
CC trimethyl)homoserine) synthetic enzymes. Bsal (and BtaA and BtaB) enzymes  
CC allow the replacement of phospholipids with non-phosphorus containing  
CC lipids in the cell membrane. A transgenic plant expressing the enzymes  
CC would have a reduced need for phosphate-containing fertilizer. Also  
CC included are an RNA transcribed from the purified DNA, antibodies  
CC produced from the protein, a vector comprising the DNA or nucleic acid, a  
CC host cell comprising the vector, transgenic plants comprising the vector,  
CC and a protein translated from the RNA or encoded by the nucleic acid. The  
CC composition is useful in producing betaine lipid compounds e.g.,  
CC diacylglycerol-O-4'-(N,N,N',N'-trimethyl) homoserine (DGTS). The composition  
CC of the invention reduces the amount of phosphate fertilizer needed for  
CC the optimal growth of crop plants. The present sequence encodes a betaine  
CC lipid synthetic enzyme.

Sequence 5242 BP; 955 A; 1562 C; 1662 G; 1063 T; 0 U; 0 Other;

Query Match 33.1%; Score 645; DB 14; Length 5242;  
Best Local Similarity 100.0%; Pred. No. 8.1e-110;  
Matches 645; Conservative 0; Mismatches 0; Indels 0

QV 1045 CAGCTGTTCCGGCAGGGCGTGACCCCGCATTTGAGAGCTGTACGAGAAGAGCTGGCG 1104

Db 3228 CAGCTGTTCCGCGAGGGCGTGCA CCGCGCATGTAGGAGCTGTACGAGAGAGAGCTGGCG 3287

1105	CCCTTCTGTGCGAAACACGACCAAACTTCTGTGTGTCGCAAGCGCCTCTGTGTAATCTCCAGCAC	1164
Db	CCCTTCTGTGCGAAACACGACCAAACTTCTGTGTGTCGCAAGCGCCTCTGTGTAATCTCCAGCAC	3347
1165	GGCCTGTACTACAGGGCGCATGGCGAAGCTGTGTGGTGTGTGAGTGCCTCTGGCCGCTG	1224
Db	GGCCTGTACTACAGGGCGCATGGCGAAGCTGTGTGGTGTGTGAGTGCCTCTGGCCGCTG	3407
1225	GTGCTGGGACTGGGCAAGACCCGTCAAGCGCCTCGCCAAACGCGCCCAAACTGGAGGAGCAG	1284
Db	GTGCTGGGACTGGGCAAGACCCGTCAAGCGCCTCGCCAAACGCGCCCAAACTGGAGGAGCAG	3467
1285	CGCGCTGTGTGGAGACGACCAATGCTATCCATTCTGTGAAGAACGGGCCCAAGCGCGCTG	1344
Db	CGCGCTGTGTGGAGACGACCAATGCTATCCATTCTGTGAAGAACGGGCCCAAGCGCGCTG	3527
1345	GTGTGGCTGTTCGTCAAGTTTCGTGAGCCTGTGTCCTTCAAAAGGCGCGTGTGTGTTC	1404
Db	GTGTGGCTGTTCGTCAAGTTTCGTGAGCCTGTGTCCTTCAAAAGGCGCGTGTGTGTTC	3587
1405	GGCGGGCGGTGCGCGGCAAGCAGTAGTCGGCTGATCAAGCGCGACGGCATCCCCATTGAG	1464
Db	GGCGGGCGGTGCGCGGCAAGCAGTAGTCGGCTGATCAAGCGCGACGGCATCCCCATTGAG	3647
1465	AATACATCGCGGCAACCATGGACGGCGTGGCGGAGAACTCGCAGTCGCGCAAGCAGAAC	1524
Db	AATACATCGCGGCAACCATGGACGGCGTGGCGGAGAACTCGCAGTCGCGCAAGCAGAAC	3707
1525	TACTTCTACTACAACTGCCTCACCGGCAAGTTCCTCGCGGACAACTGCCCCACTACCTG	1584
Db	TACTTCTACTACAACTGCCTCACCGGCAAGTTCCTCGCGGACAACTGCCCCACTACCTG	3767
1585	CGCGAGGCGGCTTCGCCACCCCTCAAGAGTGGCGTGTGGACAACTGACCGTCTCCACC	1644
Db	CGCGAGGCGGCTTCGCCACCCCTCAAGAGTGGCGTGTGGACAACTGACCGTCTCCACC	3827
1645	AATCTTCTTATGAGGAGCTCAAGCGCGCACCCTACCAACGAGTG	1689
Db	AATCTTCTTATGAGGAGCTCAAGCGCGCACCCTACCAACGAGTG	3872

RESULT 3

ABS58051

ABS58051  
ID ABS58051 standard: DNA: 1252 BP.

XX  
AC

XX  
DT 05-FEB-2003 (first entry)

XX DE Gene encoding Rhodobacter sphaeroides btaA protein.

XX Betaine lipid production; betaine lipid compound; DGTS;  
KW diacylglycerol-0-4'-(N,N,N-trimethyl) homoserine; agri  
KW phosphate-containing fertiliser; transgenic; btaA; gne  
KW

XX Rhodobacter sphaeroides. OS

XX	Key	Location/Qualifiers
FH		

FT	
FT	1. .1251
FT	/*taq= a

```

FT      /partial
FT      /product= "BtaA protein"

```

```
FT /transl_except= (pos:1. .3, aa:Met)
FT /note="This sequence lacks a start codon"
```

PN WO200283844-A2.

XX  
PD 24-OCT-2002.

09-APR-2002:

13-APR-2001: 2001US-0283812P.

\_\_\_\_\_





Db 182 TCGCATCCGCGCGGCGGCTGTGTGCGCATCGCTCGGGGGTTCGCAAGTGCTTT 241  
Qy 929 ACTGTCTGTGTCAGGCGCGCGCGAGTGTGTGTGTGTCAGTCAACCCGCGCAGTCGG 988  
Db 242 CCTATCTCACGCGGCGCGGCTCGATCTCTGCGCGTGGATCTCTGCGCGCCCATGTGG 301  
Qy 989 CGCTTCTGAGCTGTAAGAGGTGGCCATTACAGAGCTG---GAGTTGAGAGAGTGTGGC 1045  
Db 302 CGCTGGGCGGCTGAAGCTCGCGCGCGCGAGCGCTGCGCGACCATGCGCCTTCTTGG 361  
Qy 1046 AGCTGTTTCGCGAGGCGGTGCACCGCGCATTCAGGAGCTGTACGAGAAGCTGCGCG 1105  
Db 362 ATCTTCTGTGCGCAGACTGCGCGCGCATGCGGCGCTCTACGACCGCCATCGCGC 421  
Qy 1106 CTTCTGTGTCGAACACAGCCACAATCTTGTGTCGAAGCGCCTCTGTACTTCCAGCAG 1165  
Db 422 CCGCGCTCGAGCGCGCGCGCGCTACTCTGAGGCGCGCAGCGCCCTTCGCGCGCGCA 481  
Qy 1166 GCCTGTACTACGAGCGCGCATGGGCAAGCTGTGCTGGGTGCTGCACTGCTGCGCG--- 1222  
Db 482 TCCAGCTGTTCGAGCGCGGCTTCTACCGGCAAGCTGCGCGCTTCATCGCGCGG 541  
Qy 1223 -----TGGTCTGCGGACTGCGGCAAGACCGTCAAGCGCCTCGCCCAAGCGCCACAA 1273  
Db 542 CCATAGCTGCG 601  
Qy 1274 TGAGGAGCAGCGCGCGCTCTGTGGGACAGCAACATGCTCATCTTCTGTAAGAACGGCG 1333  
Db 602 TCGAGGCGCAGCGCAGCTTCTTACGCCCATATCGGCGCGCTCTCGAGCGCGCGTGG 661  
Qy 1334 CCAAGCGCGTGTGTGCTGTTGTCAGTTCGTGAGCTGAGCTGCTGCTC 1380  
Db 662 TGCAGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCTCTTCGCGCGTGGGATC 708

## RESULT 5

ADP98536  
ID ADP98536 standard; DNA; 2259 BP.

XX AC ADP98536;

XX DT 23-SEP-2004 (first entry)

XX DE C. albicans specific gene, orf6.3438, DNA sequence.

XX KW Diploid fungal cell; allele; gene disruption cassette;  
XX KW promoter replacement fragment; antifungal; fungicide; gene therapy;  
XX KW infection; Candida albicans; gene; ds.

XX OS Candida albicans.

XX PN WO2004056965-A2.

XX PD 08-JUL-2004.

XX PF 19-DEC-2003; 2003WO-US040618.

XX PR 19-DEC-2002; 2002US-0434832P.

XX XX (ELIT-) ELITRA PHARM INC.

XX PA (ELIT-) ELITRA CANADA LTD.

XX PI Roemer T, Jiang B, Boone C, Bussey H;

XX DR WPI; 2004-500296/47.

XX DR P-PSDB; ADP98846.

XX PT Constructing a strain of diploid fungal cells in which both alleles of a  
XX PT gene are modified comprises modifying the alleles of a gene in the fungal  
XX PT cells by recombination using a gene disruption cassette and a promoter  
XX PT replacement fragment.

XX PS Claim 33; SEQ ID NO 6021; 163pp; English.

XX

The invention relates to a novel method for constructing a strain of diploid fungal cells in which both alleles of a gene are modified. The method comprises modifying the alleles of a gene in diploid fungal cells by recombination using a gene disruption cassette and a promoter replacement fragment. The invention further comprises: assembling a collection of diploid fungal cells each of which comprises modified alleles of a different gene; a strain of diploid fungal cells comprising modified alleles of a gene, where the first allele of the gene is inactivated by a gene disruption cassette comprising a nucleotide sequence encoding an expressible selectable marker; and the expression of the second allele of the gene is regulated by a heterologous promoter that is operably linked to the coding region of the second allele of the gene, and where the gene encodes the polypeptide mentioned above; a collection of diploid fungal strains comprising the diploid strains cited above, where substantially all the different genes that encode the above amino acid sequences are modified and are present in different diploid strains in the collection; a nucleic acid molecule microarray comprising nucleic acid molecules, where each nucleic acid molecule comprises a nucleotide sequence that is hybridizable to a target nucleotide sequence comprising any of the 310 nucleotide sequences listed in the specification (ADP98516-ADP98825); identifying a gene that is essential to the survival or growth of a fungus, that contributes to the virulence and/or pathogenicity of a fungus, or that contributes to the resistance of a diploid fungus to an antifungal agent; identifying an antifungal agent that inhibits the growth of a diploid fungus, or a therapeutic level of proteins or gene transcripts with the inhibition of growth or proliferation of a diploid fungal cell; a purified or isolated nucleic acid molecule comprising a nucleotide sequence encoding a gene product required for proliferation of Candida albicans, where the gene product consists of any of the above-mentioned amino acid sequences; a vector comprising a promoter operably linked to the nucleic acid molecule cited above; a host cell containing the vector; a purified or isolated polypeptide comprising any of the 61 amino acid sequences given in the specification (ADP96718-ADP96779); a fusion protein comprising a fragment of a first polypeptide fused to a second polypeptide, the fragment consisting of at least 6 consecutive residues of any of ADP98826-ADP99135; producing a polypeptide; identifying a compound which modulates the activity of a gene product encoded by a nucleic acid comprising any of ADP98516-ADP98825; eliciting an immune response in an animal; a strain of Candida albicans, where a first allele of a gene comprising any of ADP98516-ADP98825 is inactive and a second allele of the gene is under the control of a heterologous promoter; identifying a compound or binding partner that binds to the polypeptide comprising any of ADP98826-ADP99135, or its fragment; identifying a compound having the ability to inhibit growth or proliferation of Candida albicans; inhibiting growth or proliferation of Candida albicans cells; manufacturing an antimycotic compound; treating an infection of a subject by Candida albicans; preventing or containing contamination of an object by Candida albicans, or for preventing or inhibiting formation on a surface of a biofilm comprising Candida albicans; a pharmaceutical composition comprising a therapeutic amount of an agent which reduces the activity or level of a gene product encoded by a nucleic acid comprising any of ADP98516-ADP98825 in a pharmaceutical carrier; an antibody preparation which binds the polypeptide; methods for evaluating a compound against a target gene product encoded by any of ADP98516-ADP98825; identifying an antimycotic compound; a computer or a computer readable medium that comprises at least one of the nucleotide sequences mentioned in the specification or at least one amino acid sequence selected from ADP98826-ADP99135; a method assisted by a computer for identifying a putatively essential gene of a fungus; and a protein array comprising proteins, where at least one protein comprises an amino acid sequence or a portion of an amino acid sequence selected from ADP98516-ADP98825. The novel methods and compositions have fungicide activity. The compositions may be used in gene therapy. The composition and methods are useful for drug screening purposes or for diagnosing, preventing or treating infections associated with Candida albicans. These may also be used for constructing strains useful for identification and validation of gene products as effective targets for therapeutic intervention, for identifying and validating gene products as effective targets for therapeutic intervention, and for collecting identified essential genes. This polynucleotide sequence represents a Candida albicans fungal specific gene of the invention.

CC NOTE: This sequence was downloaded from an electronic sequence listing  
CC provided on the WIPO website.

Seq	Sequence	2259 BP;	714 A;	359 C;	406 G;	780 T;	0 U;	0 Other;
	Query Match	3.9%;	Score 76.8;	DB 12;	Length 2259;			
	Best Local Similarity	54.3%;	Prod. No. 8.5e-05;					
	Matches	178;	Conservative	0;	Mismatches 147;	Indels 3;	Gaps 1;	
Qy	212	ACCTCACTCGGTTGAC	TGGTGGTGGCGAC	TGGGGAGAA	TGTCGATATGAT	GGCTGATT	271	
Db	323	ATCTATTGGATTGAT	TATGGTGGTGGAACT	TGGTCCAA	TATTGAATTCAT	GAATGAA	382	
Qy	272	ACATCGACCT---	GGCGAAGTTC	CAAGTCCAT	CTACGTGGT	CGACCTGT	GCCACCTCGCTGT	328
Db	383	TTAGTAAATATCT	GAAACCTTTAA	AGCTGTTT	ATTTGGTTGAT	CTTTCCCA	CTCTTTGT	442
Qy	329	GCGAGGTGGCCAA	GAAAGCGGCA	AGGGCTGGA	GAAGTGC	CAGTCTG	TGGAGG	388
Db	443	GTGAAGTTGCTTA	AGGCAAGATTT	GAAGGCCA	CTGAATGGA	CAAAATGTT	CATGATATTAGTTG	502
Qy	389	CCGACGCTGCGCA	ATTTGGCGCCCT	CTGAGGGC	ACCGGACGCT	CATACACCTT	CTTCTCCTACT	448
Db	503	CTGATGCTGTGAT	TTTACTATTGAT	TATGATAGT	GTGATTTGAT	TACTTTTCT	TATTTCTTATT	562
Qy	449	CGCTCAGCATGAT	TTCCACCGTT	CCACAACGTC	ATCGACCAGGCT	TGCTCGTAC	CTGCCCC	508
Db	563	CATTGTCGATGAT	CCCAACTTTT	CAATGCTGCT	ATCGATAAT	GTCTTCT	TAAATTAGATA	622
Qy	509	AAGACGGCCTGGT	GGGGCTTGCC	CACTT	536			
Db	623	TGGAAGGTATTAT	TGGCCACTCTG	GATTT	650			

RESULT 6	
ADW72752	
ID	ADW72752 standard; DNA; 2727 BP.
XX	
AC	ADW72752;
XX	
DT	21-APR-2005 (first entry)
XX	
DE	Neurospora crassa Btal coding region.
XX	
KW	Phospholipid synthesis; transgenic plant; fertilizer; ss; gene;
KW	betaine lipid.
XX	
OS	Neurospora crassa.
XX	
Key	Location/Qualifiers
CDS	1..2727
FT	/*tag= a
FT	/product= "Btal"
FT	
XX	
WO	2005009115-A2.
PN	
XX	
PD	03-FEB-2005.
XX	
PF	15-JUL-2004; 2004WO-US022789.
XX	
PR	16-JUL-2003; 2003US-00620914.
XX	
PA	(UNMS ) UNIV MICHIGAN STATE.
XX	
PI	Benning C, Riekhof W;
XX	
DR	WPI; 2005-112975/12.
DR	P-PSDB; ADW72753.
XX	
XX	New composition comprising a purified DNA having an oligonucleotide
PT	sequence encoding a protein, useful in producing Betaine lipid compounds
PT	e.g., Diacylglyceryl-O-4'-(N,N,N,-trimethyl) homoserine (DGTS).
XX	

PS Claim 1; SEQ ID NO 49; 147pp; English.

XX  
XX  
CC The invention relates to a new composition comprising a purified DNA  
CC having an oligonucleotide sequence appearing as ADW72747 (*Chlamydomonas*  
CC reinhardtii Btal cDNA) or ADW72752 (*Neurospora crassa* Btal coding region,  
CC encoding Btal proteins) which are DGTS (diacylglycerol-O-4'-(N,N,N-,  
CC trimethyl)homoserine) synthetic enzymes. Btal (and BtaA and BtaB) enzymes  
CC allow the replacement of phospholipids with non-phosphorus containing  
CC lipids in the cell membrane. A transgenic plant expressing the enzymes  
CC would have a reduced need for phosphate-containing fertilizer. Also  
CC included are an RNA transcribed from the purified DNA, antibodies  
CC produced from the protein, a vector comprising the DNA or nucleic acid, a  
CC host cell comprising the vector, transgenic plants comprising the vector,  
CC and a protein translated from the RNA or encoded by the nucleic acid. The  
CC composition is useful in producing Betaine lipid compounds e.g.,  
CC Diacylglycerol-O-4'-(N,N,N-,trimethyl) homoserine (DGTS). The composition  
CC of the invention reduces the amount of phosphate fertilizer needed for  
CC the optimal growth of crop plants. The present sequence encodes a betaine  
CC lipid synthetic enzyme.

XX  
XX  
CC Sequence 2727 BP: 616 A; 766 C; 713 G; 632 T; 0 U; 0 Other

Query Match	3.7%	Score 73	DB 14	Length 2727	
Best Local Similarity	51.0%	Pred. No. 0.00044			
Matches	172	Conservative 0	Mismatches 165	Indels 0	Gaps 0
Qy	779	ACGAGGAGCGTGGAGCGCGCCGCCCATGTTCCGCCCCACCTTCCTGTACACGACGTCGT	838		
Db	1448	ACGACGATCAACTCCGGAAGCA CACCAGTTCNAATGACGAGTACATCTACGCTTTACCT	1507		
Qy	839	GGGAGGACCCCGAGCGCGATATGGAGGTGATGGAGATCAACCCCAAGGACACGGTGCTGA	898		
Db	1508	GGGAAGACTCGCGGTGCACAGAACTCTTAACTCGGGCCCGACGAGCTGCTCTAG	1567		
Qy	899	CCCTGACTAGCGCGGTGCAATGCCCTGAACCTGTGTGCAGGGGGCGCGCCAGGTGG	958		
Db	1568	CCATCACACGCGCGGCGACAACATTCTTTCTTACCTGATGCAGAGTCCCGCTCGCGTGC	1627		
Qy	959	TGTCGTGGACTGCACACCCCGCGAGTCGCGGCTTCTGGAGCTGAGAAAGTGGCCATTC	1018		
Db	1628	ACGCCATCGACCTAAACCCAGCCCAAACACCTGCTTGAACCTCAAAGTCGCCTCTTTTA	1687		
Qy	1019	AGCAGCTGGAGTTTGGAGAGCTGTGGCAGCTGTTTCGGCAGGGCGTGCACCCCGCGCATG	1078		
Db	1688	CGACTCTGGATTACCCCGACGCTCTGGNAGATCTTCGTGAGGGCAACACCCCGACTTTC	1747		
Qy	1079	AGGAGCTGTACGAGAAGACTGGCGCCCTTCCTGTGTC	1115		
Db	1748	GCTCACTGCTCATCTCCAAACTCTCCCTCAGCTCTC	1784		

RESULT 7	
ADW72751	
ID	ADW72751 standard; DNA; 3427 BP.
XX	
XX	
AC	
AC	ADW72751;
XX	
DT	21-APR-2005 (first entry)
XX	
XX	Neurospora crassa Btal gene.
DE	
XX	
XX	Phospholipid synthesis; transgenic plant; fertilizer; ds; gene;
KW	betaine lipid.
KW	
XX	
OS	Neurospora crassa.
XX	
XX	
FN	WO2005009115-A2.
XX	
XX	
PD	03-FEB-2005.
XX	
XX	
PF	15-JUL-2004; 2004WO-US022789.
XX	
XX	
PR	16-JUL-2003; 2003US-00620914
PR	



```
FT      CDS      encoded protein is not provided"
FT      CDS      complement(8070..9989)
FT      CDS      /*tag= l
FT      CDS      /product= "Parapoxvirus ovis polypeptide"
FT      CDS      /note= "This ORF is specifically claimed within claim 3
FT      CDS      of the specification, although the sequence of the
FT      CDS      encoded protein is not provided"
FT      CDS      complement(10062..11195)
FT      CDS      /*tag= m
FT      CDS      /product= "Parapoxvirus ovis polypeptide"
FT      CDS      /note= "This ORF is specifically claimed within claim 3
FT      CDS      of the specification, although the sequence of the
FT      CDS      encoded protein is not provided"
FT      CDS      11227..11493
FT      CDS      /*tag= n
FT      CDS      /product= "Parapoxvirus ovis polypeptide"
FT      CDS      /note= "This ORF is specifically claimed within claim 3
FT      CDS      of the specification, although the sequence of the
FT      CDS      encoded protein is not provided"
FT      CDS      11802..12038
FT      CDS      /*tag= o
FT      CDS      /product= "Parapoxvirus ovis polypeptide"
FT      CDS      /note= "This ORF is specifically claimed within claim 3
FT      CDS      of the specification, although the sequence of the
FT      CDS      encoded protein is not provided"
FT      CDS      complement(12080..12358)
FT      CDS      /*tag= p
FT      CDS      /product= "Parapoxvirus ovis polypeptide"
FT      CDS      /note= "This ORF is specifically claimed within claim 3
FT      CDS      of the specification, although the sequence of the
FT      CDS      encoded protein is not provided"
FT      CDS      complement(12364..13980)
FT      CDS      /*tag= q
FT      CDS      /product= "Parapoxvirus ovis polypeptide"
FT      CDS      /note= "This ORF is specifically claimed within claim 3
FT      CDS      of the specification, although the sequence of the
FT      CDS      encoded protein is not provided"
FT      CDS      complement(14053..14826)
FT      CDS      /*tag= r
FT      CDS      /product= "Parapoxvirus ovis polypeptide"
FT      CDS      /note= "This ORF is specifically claimed within claim 3
FT      CDS      of the specification, although the sequence of the
FT      CDS      encoded protein is not provided"
FT      CDS      15080..15394
FT      CDS      /*tag= s
FT      CDS      /product= "Parapoxvirus ovis polypeptide"
FT      CDS      /note= "This ORF is specifically claimed within claim 3
FT      CDS      of the specification, although the sequence of the
FT      CDS      encoded protein is not provided"
FT      CDS      complement(15423..16838)
FT      CDS      /*tag= t
FT      CDS      /product= "Parapoxvirus ovis polypeptide"
FT      CDS      /note= "This ORF is specifically claimed within claim 3
FT      CDS      of the specification, although the sequence of the
FT      CDS      encoded protein is not provided"
FT      CDS      complement(16847..19021)
FT      CDS      /*tag= u
FT      CDS      /product= "Parapoxvirus ovis polypeptide"
FT      CDS      /note= "This ORF is specifically claimed within claim 3
FT      CDS      of the specification, although the sequence of the
FT      CDS      encoded protein is not provided"
FT      CDS      19156..19704
FT      CDS      /*tag= v
FT      CDS      /product= "Parapoxvirus ovis polypeptide"
FT      CDS      /note= "This ORF is specifically claimed within claim 3
FT      CDS      of the specification, although the sequence of the
FT      CDS      encoded protein is not provided"
FT      CDS      complement(19736..20314)
FT      CDS      /*tag= w
FT      CDS      /product= "Parapoxvirus ovis polypeptide"
FT      CDS      /note= "This ORF is specifically claimed within claim 3
FT      CDS      of the specification, although the sequence of the
FT      CDS      encoded protein is not provided"

10201..22101
/*tag= x
/product= "Parapoxvirus ovis polypeptide"
/note= "This ORF is specifically claimed within claim 3
of the specification, although the sequence of the
encoded protein is not provided"
22125..22940
/*tag= y
/product= "Parapoxvirus ovis polypeptide"
/note= "This ORF is specifically claimed within claim 3
of the specification, although the sequence of the
encoded protein is not provided"
23003..23866
/*tag= z
/product= "Parapoxvirus ovis polypeptide"
/note= "This ORF is specifically claimed within claim 3
of the specification, although the sequence of the
encoded protein is not provided"
complement(23873..26908)
/*tag= aa
/product= "Parapoxvirus ovis polypeptide"
/note= "This ORF is specifically claimed within claim 3
of the specification, although the sequence of the
encoded protein is not provided"
26926..27213
/*tag= ab
/product= "Parapoxvirus ovis polypeptide"
/note= "This ORF is specifically claimed within claim 3
of the specification, although the sequence of the
encoded protein is not provided"
complement(27216..27626)
/*tag= ac
/product= "Parapoxvirus ovis polypeptide"
/note= "This ORF is specifically claimed within claim 3
of the specification, although the sequence of the
encoded protein is not provided"
complement(27616..29754)
/*tag= ad
/product= "Parapoxvirus ovis polypeptide"
/note= "This ORF is specifically claimed within claim 3
of the specification, although the sequence of the
encoded protein is not provided"
complement(29800..32217)
/*tag= ae
/product= "Parapoxvirus ovis polypeptide"
/note= "This ORF is specifically claimed within claim 3
of the specification, although the sequence of the
encoded protein is not provided"

Query Match      3.7%; Score 72.4; DB 12; Length 137560;
Best Local Similarity 42.9%; Pred. No. 0.00081;
Matches 461; Conservative 0; Mismatches 611; Indels 2; Gaps 2;

618  CGACAACTTGACATCGGCCCGCGCCGCTTACCTGAGCAGAGCTGGAGCGCGT 677
      |||||
20613 CGAGGTCTTCGAGCGCGCTGCTCAAGGCGCGCTTCGAGCGCCCTAGCTCTTTGTCGGGC 20672
      |||||
678  GTGGGAGCAGAACACCCAGGGTTTCGATCCCTACGTCCCGTGGCTGCGCCCTACTA 737
      |||||
20673 GCTCTCGGCGCATCTTCAGGCGCTTCGTCGCGCATCCGGGCTACCGCGGACCTGCGC 20732
      |||||
738  CGTGTGGATTTGGCGCGCTGCCAGCGTTGGCCAGCCCTTCGACGAGAGCGCGTGGAGCG 797
      |||||
20733 CGAGCTGCTGCGCGACGCGCGCTTCATGAGAGTGGCTGCGCGCGCGCGGAGCTCGCGGA 20792
      |||||
798  GCGGCCCATGTTCCCGCCACCTTCCTGTACAGCAGCTGCTGGGAGACCCCGAGCGGA 857
      |||||
20793 GGTCAATTTGGCGTGAACACGACATTCGCGCGAACCCGCTCTTCGCGAGCGGAGCCCGT 20852
      |||||
858  TATGGAGGTGATGAGATCAACCCCGAGGACACGGTGTCTGACCCCTGACTAGCGCGCGCTG 917
      |||||
20853 GCGCGACGCGGAGCTCATTTTCGCAAGACCTTCGCAAGACCGAGTTCGCGCGCGTCAA 20912
      |||||
918  CAATGCCCTGAACCTGCTGTCGAGGGGCGCGCGAGTGGTGTCTGCTGAGTCTGCAACCC 977
      |||||
```







Db 251 CGCGGACGAGATCGAGCGCGTGCACCTCAACGCGCGCCCAATCGCGCTGAACCGCATGA 310  
Qy 1004 AGAAGTGGCGCAATTCAGCAGCTGGAGTTTGAGGACG---TGTGGCAGCTCTTCGGCGAGG 1060  
Db 311 AGCTGGAGGCGGTGGCGCTGCTCCCTCGCAGGCGGATCTGTTCCGCTTTTTCGGCGCG 370  
Qy 1061 GCGTGCAACCGCGCATTTGAGAGGTGTACAGAGAAAGCTGGCGGCCCTCTCTGCGCAAA 1120  
Db 371 CGACACACGAGCCCAATTCGCAAGCCCTATGACCGCTTTATTTGGCGCGCATCTCGATCCGG 430  
Qy 1121 CCAGCCCAAACTTCTGTCTCAAGCGCTCTGTGTAATTCACAGCAGCGCTGTACTACCAAG 1180  
Db 431 TCAGCCGCCACTATTGGAGCGCGCAACTGGCGTGTCCGCGGCAATCGCCGTCTTCG 490  
Qy 1181 CGCGCATGGCAAGCTGTGTGGTGTCTGCAGTGCCTCGGCGCGTGTGTCTGGAGCTGGCA 1240  
Db 491 ACCGCAATTTTACCAGAC---GGCCTGTCTGCGCTGTTTCATCGCCATGGCCATCGCA 547  
Qy 1241 AGACCGTCAAGCGCTCGCAACGCGCCCAATGAGGAGCAGCGCGCTGTCTGGGACA 1300  
Db 548 CGCGCAAAATTTCTTGGCGTCAACCGCGCCCAATGATGGAAGCAGCAATATCGCGGAGC 607  
Qy 1301 GCAACATGCTCATCTCTGTGTAAGAACGCGCCCAAGCGCTGTGTGTCTGTCTGTC 1360  
Db 608 AGCGCGCTTCTTCAACGAG-----AGCTGGCGCGGTCTTCGACAAGA 652  
Qy 1361 AGTTCTGAGCGCTGTGTCTTCAACAAGCGCGTGTGTGTGTGGCGCGCGGTGCGCG 1420  
Db 653 AGCTTTTGAATGGCGACCTCGGTAAAGCGCTCGCTGTTCGGCTCGGCATTCGCGCG 712  
Qy 1421 GCAAGCATGAGCGCTGATCAAGCGGAGCGGATCCCATTTGAGAACTATCATCGCGGCA 1480  
Db 713 CGCAGTACGATTCCTGATCACCTCAGCGCAGCGCACCATGGCGCAGCGTTCTGAAGGCC 772  
Qy 1481 CCATGACGCGGTGGCGGAATCTCGCAGCTGCGCAGCAGCACTACTTCTACTCAACT 1540  
Db 773 GCGTGGAAAGCTCGCTGCGATTTTCCCTTGGAAACAATTTATTCGCTTGGCGGCTT 832  
Qy 1541 GCCTCACCGGCAAGTTCTCTCGCGCAACAATGCG-----CCCACTACTCGCGGAGCGG 1594  
Db 833 TTGCCCGCGCTATCCAAATCCCGGTGAGCGCGCTGCGCGCTATCTGGAAGCAGA 892  
Qy 1595 CTTTGGCACCTT-----CAAGATGGCGGTGTGGAACAACCTGACGCTTCCACCACT 1648  
Db 893 ACTACGAAACCACTCGCGGCAATATCGACGCGTGGCCATCCACCATGCAATCTGATCG 952  
Qy 1649 TCTTATGAGGAGCTCAAGCGCGCACCTACACCAAGGTGATTCATGACCACTGCG 1708  
Db 953 AATTCCTGCGCGGAGGACGCGGCACTCAACGCGTGTGGTGGAAATCAGCGCACCGCTCG 1012  
Qy 1709 ACTGGCTGGATATGCCGTGGCCAAAGAGCTGGCGGAGTGCCTGGCCAAAGAGTTGGCG 1768  
Db 1013 ACTGGATGACGATGACCACTCAACGCGTGTGGTGGAAATCAGCGCACCGCTCG 1072  
Qy 1769 CGGGCGCATGTCATCTGGCGCTCGCGCTCC 1800  
Db 1073 CAGCGCGCGCTCATCTTCGCAACGCGCGCC 1104

## RESULT 12

ADW72725  
ID ADW72725 standard; DNA; 1248 BP.

XX

AC ADW72725;

XX

DT 21-APR-2005 (first entry)

XX

DE Mesorhizobium loti btaA gene.

XX

KW Phospholipid synthesis; transgenic plant; fertilizer; ss; gene;

XX

XX betaine lipid.

XX

OS Mesorhizobium loti.

XX Key Location/Qualifiers  
FH CDS 1. .1248  
PT /\*tag= a  
FT /product= "BtaA protein"  
XX

PN W02005009115-A2.

XX

PD 03-FEB-2005.

XX

PP 15-JUL-2004; 2004MO-US022789.

XX

PR 16-JUL-2003; 2003US-00620914.

XX

PA (UNMS ) UNIV MICHIGAN STATE.

XX

PI Benning C, Riekhof W;

XX

XX WPI; 2005-112975/12.

DR P-PSDB; ADW72744.

XX

PT New composition comprising a purified DNA having an oligonucleotide sequence encoding a protein, useful in producing Betaine lipid compounds e.g., Diacylglycerol-O-4'-(N,N,N-trimethyl) homoserine (DGTS).

XX

PS Disclosure; SEQ ID NO 22; 147pp; English.

XX

CC The invention relates to a new composition comprising a purified DNA having an oligonucleotide sequence appearing as ADW72747 (Chlamydomonas reinhardtii Bta1 cDNA) or ADW72752 (Neurospora crassa Bta1 coding region, encoding Bta1 proteins which are DGTS (diacylglycerol-O-4'-(N,N,N-trimethyl)homoserine) synthetic enzymes. Bta1 (and BtaA and BtaB) enzymes allow the replacement of phospholipids with non-phosphorus containing lipids in the cell membrane. A transgenic plant expressing the enzymes would have a reduced need for phosphate-containing fertilizer. Also included are an RNA transcribed from the purified DNA, antibodies CC produced from the protein, a vector comprising the DNA or nucleic acid, a host cell comprising the vector, transgenic plants comprising the vector, CC and a protein translated from the RNA or encoded by the nucleic acid. The CC composition is useful in producing Betaine lipid compounds e.g., CC diacylglycerol-O-4'-(N,N,N-trimethyl) homoserine (DGTS). The composition CC of the invention reduces the amount of phosphate fertilizer needed for CC the optimal growth of crop plants. The present sequence encodes a betaine CC lipid synthetic enzyme.

XX

SQ Sequence 1248 BP; 224 A; 441 C; 345 G; 238 T; 0 U; 0 Other;

Query Match 3.6%; Score 69.6; DB 14; Length 1248;

Best Local Similarity 45.0%; Pred. No. 0.0017;

Matches 500; Conservative 0; Mismatches 579; Indels 33; Gaps 5;

Qy 704 TCCCTACGTGCGTGGCTGCGCGCCCTTACTACGTGTGTGATGCGCGCTGCCACGCG 763

Db 11 TCTCCTCGATCTGGTTTTCGCGCGGCAAGAGTTGGAAGGCGCTTACCAAGACC 70

Qy 764 TTGGCCACGCCCTGCAAGAGGAGCGGTGAGCGCGCGCCATGTTCCCGCCACCTTCC 823

Db 71 GCGCGCTTTCCAAAGCGCGCATCTCGAGCGGCTTTCGCTTCTGTTTTCGCGCTCG 130

Qy 824 TGTACAGCAGTCTGGAGGAGCCCGAGCCGATATGAGGTGATGAGATCAACCCCA 883

Db 131 TCTATCCGAGATCTGGGAAGACCCCGATGTGACATGAGGCGCATGCGCTTGGTCAGG 190

Qy 884 AGGACACGCTGTGACCTGACTAGCGCGGCTTGCATGCGCTTGAACCTGCTGTGTGAGG 943

Db 191 GCCATCGCATGTCACAAATGCTTTCGCGGCTTGCACATCTCTGCGCTTACCTACCGGTT 250

Qy 944 GGGCCCGCCAGGTGGTGTGCGTGGACTGCAACCCCGCGAGTCGGCGCTTCTGGAGCTGA 1003

Db 251 CGCGCGCAGGATCGAGCGCGTGCACCTCAACGCGCCCAACATCGCGCTGAACCGCATGA 310

Qy 1004 AGAAGTGGCATTTACGACGCTGGAGTTTGAAGAGC---TGTGGCAGCTGTTGGCGAGG 1060

```
Db 311 AGCTGGAGGCGTGGCGCTGTGCTCGCAGGCGGATCTGTTCGCTTTTTCGGCGCG 370
Qy 1061 GGTGACACCGCGCATTTGAGAGCTGTACAGAGAGCTGGCGCCCTTCTCTGCGCAA 1120
Db 371 CCAGACACAGCCCAATTCAGAGCCTATGACCGCTTTATTTGGCGCGCATCTCGATCCGG 430
Qy 1121 CCAGCCACAATTTCTGTCTCAAGCGCTCTGTGTACTTCCAGCAGCGGCTGTACTACCAGG 1180
Db 431 TCAGCCGCCACTATTTGGAGCGCGCAACTGGGTGTGCTGCCGGCGCATCGCGTCTTCG 490
Qy 1181 CGCGCATGGCAAGCTGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTG 1240
Db 491 ACCGCAATTTCTACGAGC---GGCCTGTGCGGCTGTGCTGTGCTGTGCTGTGCTGTG 547
Qy 1241 AGACCGTCAAGCGCTCGCAAGCGCGCCACATGAGGAGCAGCGCGCTGTGTGGACA 1300
Db 548 CGCGCAATTTCTTTCGGCGTCAACCGCGCCACATGATGGAAGCAGCAATATCGCGGAGC 607
Qy 1301 GCAACATGCTCATCCACTTGTGAAGAACGGGCCCAAGCGCTGTGTGTGCTGTGTGCTA 1360
Db 608 AGCGCCGCTTCTTCAACGAGG-----AGCTGGCGCGGCTTTCGACAGAGA 652
Qy 1361 AGTTCTGTGAGCGTGTGCTCTTCAACAAAGCGCGTGTGTGTGCTGTGCGCGCGGTGCGG 1420
Db 653 AGCTTTTGAATGGCGACCTCGCGTAAGCGCTCGCTGTTCGGCTCGGCAATTCGCGCGG 712
Qy 1421 GCAGAGTACGCGCTGATCAAGCGGACGGCATCCCATTTGAGAACTACATCGCGGCA 1480
Db 713 CGCAGTACGATTCCTGATCACTCAGCGCAGCGCACCATGCGCGCGGTTCTGAAGGCC 772
Qy 1481 CCATGGAGCGGTGGCGGAGACTCGCACGTGCGCAAGCAGAACTACTTCTACTACAAC 1540
Db 773 GGCTGGAAAGCTCGCTCGGATTTTCCCTGGAAACAATATTTTCGCTGGCAGGCTT 832
Qy 1541 GCCTCAGCGGCAAGTTCCTTGGCGCAAACTGC-----CCACCTACTCTGCGAGCGG 1594
Db 833 TTGCGCGCGCTATCCAAATCCCGGTGAGCGCGCGCTGCGCGCTATCTGGAAGCAGA 892
Qy 1595 CTTTGGCACCT-----CAAGAGTGGGTGTGGACAACTGACCGTCTCCACCACT 1648
Db 893 ACTACGAAACCATCCGCGGCAATATCGACCGGTGCGCATCCACCATGCAATCTGATCG 952
Qy 1649 TCTTCATGAGGAGCTCAAGCGCGCACCTACACCAAGGTGATCTGATGACCACTGG 1708
Db 953 AATTCTTCCGCGCAAGACGCGGCACTCGTGTGATCTCTCTCTCTCTCTCTCTCTCT 1012
Qy 1709 ACTGGCTGATATGCCCGTGGCCCAAGAGTGGCGGAGTGTGCTGGCCCAAGCAGTTCCGC 1768
Db 1013 ACTGATGACCGATGACCACTCAACGCGTGTGTCGGAATCAGCCGACCGCCTCCG 1072
Qy 1769 CGGCGCGCATGCTATCTGGGCGTTCGCGTCC 1800
Db 1073 CAGCGCGCGGTCTATCTTCGCAACCGCGCC 1104
```

## RESULT 13

AAH4047

ID AAH4047 standard; DNA; 1092 BP.

XX

AC

XX

DT 10-SEP-2001 (first entry)

XX Streptomyces sp. CL190 mevalonate pathway orfD DNA SEQ ID NO:5.

XX Streptomyces sp. CL190; mevalonate pathway; actinomycete; cardiant;

XX isoprenoid compound; osteopathic; cytostatic; ubiquitine; vitamin K2;

XX carotenoid; heart disease; osteoporosis; cancer; drug; health food; ds.

XX Streptomyces sp. CL190.

OS

XX Key Location/Qualifiers

XX FH 1..1092

XX CDS

/\*tag= a  
/product= "orfD"

WO200142476-A1.

14-JUN-2001.

06-DEC-2000; 2000WO-JP008620.

08-DEC-1999; 99JP-00348375.

(SETO/) SETO H.

(KUZU/) KUZUYAMA T.

Seto H, Kuzuyama T, Takahashi S, Takagi M;

WPI; 2001-381696/40.

P-PSDB; AAB99729.

Actinomycetes-originated genes of enzymes participating in mevalonate pathway, applicable in producing e.g. ubiquitine, vitamin K2 and carotenoids for treatment of heart diseases, osteoporosis and cancer in drug and health food.

Claim 4; Page 50-52; 75pp; Japanese.

The sequence given in AAH4043 represents a DNA sequence isolated from Streptomyces sp. CL190, containing a 6798 base pairs (SI), which encodes the whole enzyme necessary for functioning the mevalonate pathway. The sequence encodes protein sequences, designated orfA to E and hmgr, which are used in the mevalonate pathway. The proteins and polynucleotide sequences encoding them have cardiant, osteopathic and cytostatic activities. The genes are applicable in producing e.g. ubiquitine, vitamin K2 and carotenoids which can be used in the treatment of heart diseases, osteoporosis and cancer in drugs and health foods. The present sequence encodes the orfD protein from the present invention

SQ Sequence 1092 BP; 181 A; 430 C; 340 G; 141 T; 0 U; 0 Other;

Query Match 3.6%; Score 69.4; DB 4; Length 1092;

Best Local Similarity 45.3%; Pred. No. 0.0019;

Matches 253; Conservative 0; Mismatches 306; Indels 0; Gaps 0;

Qy 1344 GGTGTGCTCTTCGTCAAGTTCTGTAGCCTGTGTCTTCAACAAGCCGCTGTGTGTT 1403

Db 501 GTTCCGCGATCGAAGATCGCGCGCGCTCGACATCCCGTGATCTCAAGAGGT 560

Qy 1404 CGCGCGCGCGTCCCGGCAAGCAGTACGCGCTGATCAAGCGGAGCGGATCCCATTTGA 1463

Db 561 CGGCAACGGCTGAGCGCGGAGACCATCTGTCTGCTCGCGAGCTCGCGCTCCAGGCGGC 620

Qy 1464 GAATACATCGCGCGCACATGAGCGCGTGGCGGAGACTCGCATGTGCGAAGCAGAA 1523

Db 621 GGACGTGAGCGCGCGCGGCGGACGAGCTTCGCGCGCATCGAAGCGCGCGGAGCT 680

Qy 1524 CTACTTCTACTACAACTGCTCACCAGGAGTTCTCGCGCAAACTGCGCCCACTACCT 1583

Db 681 CGCGGACTACCGTTCTGTGACGCGTGGGGGAGTCCACCGCGCTGCTGTGGAAGC 740

Qy 1584 GCAGGAGCGCGCTTCCCGCACCTTCAAGAGTGGCGGTGGAGCAAACTGACCGTCTCCAC 1643

Db 741 CCAGGACATCTCCTGCTCCCGCTCTCGCTCCGCGGTGTGCTCACCGCTCGACGTGGT 800

Qy 1644 CAACTTCTTATGAGGAGGCTCAAGCGCGCATACACAAAGTGATTTCTGATGGACCA 1703

Db 801 CGCGCGCTCGCGTCCGCGCGCGCGCGCTGCGCTCTCTCCGCGGCTTCTTCCGCACT 860

Qy 1704 CGTGAATCTGCTGATATGCTGCGCAACGAGCTGGCGAGTGGCGTGGCGTGGCGAGGT 1763

Db 861 GATGAGCAGCGGTGCGACGCGTGTATCAGAGCTCACGACCTGGCTGGACGAGTGGC 920

Qy 1764 TGCGCGCGCGGCGCATGCTATCTGCGCTCTCGCTCTCCCTCAGCCCGCGCTACGCCGAGCT 1823

Db 921 GCGCTGCGACCATGCTCGCGCGCGCGCACCCCGCGCGACCTCAACCGCGCTGCGACGTGCT 980  
 QY 1824 GATCCGAAGCGGGCTTCGACGTGCGCTGCATCCGCGCGCCACTCAGGGCTACATGGA 1883  
 Db 981 GCTCCAGCGGAGCTGCGTGAATCTTCTGCGCGGACCGGGGATCGACACGCGCGCTCGC 1040  
 QY 1884 CCGGTCAACATGTACAGC 1902  
 Db 1041 CCAGCGTCCAGTCCATC 1059

## RESULT 14

ABD16186

ID ABD16186 standard; DNA; 3003 BP.

XX AC ABD16186;

DT 29-JUL-2004 (first entry)

XX Pseudomonas aeruginosa polynucleotide #14790.

XX Bacterial infection; gene; db; Pseudomonas aeruginosa infection;  
 KW antibacterial.

XX OS Pseudomonas aeruginosa.

XX US6551795-B1.

XX 22-APR-2003.

XX 18-FEB-1999; 99US-00252991.

XX 19-FEB-1998; 98US-0074788P.

PR 27-JUL-1998; 98US-0094190P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Rubenfield MJ, Nolling J, Deloughery C, Bush D;

XX WPI; 2003-615309/58.

DR P-PSDB; ABO82615.

PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
 useful as molecular targets for diagnostics, prophylaxis and treatment of  
 PT pathological conditions resulting from bacterial infection.

XX Disclosure; SEQ ID NO 14790; 455pp; English.

XX The invention relates to Pseudomonas aeruginosa polypeptides and the  
 CC polynucleotides encoding them. The sequences are useful in diagnosis and  
 CC therapy of pathological conditions, as molecular targets for diagnostics,  
 CC prophylaxis and treatment of pathological conditions resulting from a  
 CC bacterial infection, for evaluating a compound, such as a polypeptide,  
 CC for the ability to bind a P. aeruginosa nucleic acid, as components of  
 CC effective antibacterial targets, as targets for antibacterial drugs,  
 CC including anti-P. aeruginosa drugs, as templates for recombinant  
 CC production of P. aeruginosa-derived peptides or polypeptides, as target  
 CC components for diagnosis and/or treatment of P. aeruginosa-caused  
 CC infection, and in detection of P. aeruginosa sequences or other sequences  
 CC of Pseudomonas species using biochip technology. Sequences ABD01397-  
 CC ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:  
 CC The sequence data for this patent did not form part of the printed  
 CC specification but was obtained in electronic format from USPTO at  
 CC seqdata.uspto.gov/sequence.html

XX Sequence 3003 BP; 488 A; 1076 C; 959 G; 480 T; 0 U; 0 Other;

Query Match 3.6%; Score 69.4; DB 11; Length 3003;

Best Local Similarity 43.8%; Pred. No. 0.002;

Matches 454; Conservative 0; Mismatches 571; Indels 12; Gaps 3;

QY 521 TGGGCGTTGCGGACTTCTAGTGAGCGGCGGACGTACGACCTGCGCCCTGCGCCAGATGCCCT 580

RESULT 15

ABD15936

Db 1965 TCGGCGTTCGACGGCAGCAACCCCGACCCCTCAACACCGCGCGCCCTGTGTGATCAACCTCAAGC 2024  
 QY 581 GGTGCGCGCGTTTCTTCTGCGCATCGATCTTTCGACATTCGAACAATTCGATCGGCCCCG 640  
 Db 2025 CGCACAGCAGCGCGACGTCACCGCCAGCGAGTGATCCAGCGCTGCAGCCCGCAACTCG 2084  
 QY 641 AGCGCCGCGCTTACCTGGAGCAGAGCTGGAGCGCGTGTGGAGCAGAGAACCCAGGCTT 700  
 Db 2085 ACCACCTGCCCGGGATCAAGCTGTATGTCAGCGCGGTGTCAGGACCTGACCATCGAGACC 2144  
 QY 701 CGATCCCTACGTGCGGTGGCTGCGCGCCCTACTACGTGTGTGATTTGGCGCGCTGCCCA 760  
 Db 2145 GGTGCGCGCGACCCAGTACAGTTTACCTTGCAGGACGCCGACCGGAGCTGCTCGCG 2204  
 QY 761 GCGTTGGCCACGCCCTTCGACGAGAGCGGTGGAGCGCGGCCCTATGTTCCCGCCCACT 820  
 Db 2205 AGTGGGTGCGGAGCTGGTGGCGCGCTGCAGGAGTTGCGCGAGCTCGCCGACGTGCCA 2264  
 QY 821 TCCTGTACAGCAGTGTGGAGGACCCCGAGCCGGATATGGAGGTGATGGAGATCAACC 880  
 Db 2265 GCGACTGGCAGGACAAAGGGCTTGCAGGCTTACCTGAACATCGACCGCG-----ACACCG 2318  
 QY 881 CCAAGGACACGCTGCTGACCTGACCTAGCGCGCGCTGCAATGCCCTGAACCTGTGTGTGC 940  
 Db 2319 CTTGCGCGCTTGGCGTGAAGCTCTCCGACATTCGACAGCGTGTCTTACACGCCCTTCGCC 2378  
 QY 941 AGGGGCGCGCCAGGTGTGTGCTGCGTGGACTTCGAAACCCCGCGAGTCGCGCGCTTCTGAGC 1000  
 Db 2379 AGCGGTGATCTCGACCATCTTTCACCCAGGCCACCCAGTACCGCTGGTCTCGAGGTGG 2438  
 QY 1001 TGAAGAAGTGGCCATTTCAGCAGCTGGAGTTTTCAGGACGTGTGGAGCTGTGTGGCGAGG 1060  
 Db 2439 CGCGCAGTTTCCAGCTCGGCGCGCGAGCCCTTGAGCAGCTCTACGTGCGCTCCAGCGACG 2498  
 QY 1061 GCGTGACCCCGGCGCATTTGAGGAGCTGTACGAGAAGAGCTGCGCGCCCTCTCTGCGCAA 1120  
 Db 2499 GCACCGAGTGTGCGCTGTGAGCTGGCGAGAGTGGAGGAGCGGATACCTCTGTCGCGA 2558  
 QY 1121 CCAGCCACAACTTCTGGTCCAAAGCGCTCTGTGTAATTTCCAGCAGCGGCTGTACTACAGG 1180  
 Db 2559 TCAACCATATCGCCAGTTTC---CCCTCGCGGACCCCTGTCTTCAACCTTGGCCAAAGGTT 2615  
 QY 1181 GCGGCATGGGCAAGCTGTGTGGTGTCTGTGAGTGTCTGCGCGCTGTGTGTGTGGAGCTGGCA 1240  
 Db 2616 ACTCCCTGGCGAGCGCGTTCGAGCGGATCGTGTGCGGTTCGAGGCGAGCTGTGAGCTGCCGC 2675  
 QY 1241 AGACCGTCAAGCGCTTCGCCAACCGCCCAATGAGGAGCAGCGCGTCTGTGGGACA 1300  
 Db 2676 TGAGCATGAGGCGAGCTTTCGCGCGCGCGCGCTGTGCTTCGAGGCGCTCGCTGTGNAACA 2735  
 QY 1301 GCAACATGTCTATCC---ACTTCTGTAAGAAACGGGCGCCCAAGCGCTGTGTGTGTGTGTCG 1357  
 Db 2736 CGCTGTGTGTATCTCTCGCTCGGTGTGTGACCATGTATCATCTGCTTGGGATCTCTTACG 2795  
 QY 1358 TCAAGTTCTGAGCTGTGTGCTTTTCAACAGGCGCTGTGTGTGTGTGTGGCGGGCGGTGC 1417  
 Db 2796 AGAGCTTCTATCCATCGGTGACCATCTCTCGAACCTTGCCTCGCGCGGGGTGTGGCGCGC 2855  
 QY 1418 CCGGCAAGCAGTACCGCTGATCAAGCGGAGCGGATCCCATTTGAGAACTACATCGCGC 1477  
 Db 2856 TGTGCGCGTGTATGCTGTGGCGGGGCGAGAGATCGGCATCTGTGGGATCATCGGATATCC 2915  
 QY 1478 GCACCATGACGCGGTGTGGCGGAGAACTCGCACGTGCGCAAGCAGAACTACTTCTACTACA 1537  
 Db 2916 TGCTGATCGCATCGTCAAGAGAACCGCATCATGATGATGCTGACTTCCGCTTCGACGCG 2975  
 QY 1538 ACTGCTCTACCGGCAAG 1554  
 Db 2976 AGCGCAACGAAGGCAAG 2992

ID ABD15936 standard; DNA; 3132 BP.  
 AC ABD15936;  
 XX  
 XX 29-JUL-2004 (first entry)  
 DT  
 XX Pseudomonas aeruginosa polynucleotide #14540.  
 DE  
 XX Bacterial infection; gene; db; Pseudomonas aeruginosa infection;  
 KW antibacterial.  
 KW Pseudomonas aeruginosa.  
 OS  
 XX US6551795-B1.  
 XX  
 XX 22-APR-2003.  
 PD  
 XX 18-FEB-1999; 99US-00252991.  
 XX  
 XX 18-FEB-1998; 98US-0074788P.  
 PR  
 XX 27-JUL-1998; 98US-0094190P.  
 PR  
 XX (GENO-) GENOME THERAPEUTICS CORP.  
 PA  
 XX Rubenfield MJ, Nollong J, Deloughery C, Bush D;  
 XX WPI; 2003-615309/58.  
 XX P-PSDB; ABO82365.  
 DR  
 XX Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
 PT useful as molecular targets for diagnostics, prophylaxis and treatment of  
 PT pathological conditions resulting from bacterial infection.  
 PT  
 XX Disclosure; SEQ ID NO 14540; 455pp; English.  
 PS  
 XX The invention relates to Pseudomonas aeruginosa polypeptides and the  
 CC polynucleotides encoding them. The sequences are useful in diagnosis and  
 CC therapy of pathological conditions, as molecular targets for diagnostics,  
 CC prophylaxis and treatment of pathological conditions resulting from a  
 CC bacterial infection, for evaluating a compound, such as a polypeptide,  
 CC for the ability to bind a P. aeruginosa nucleic acid, as components of  
 CC effective antibacterial targets, as targets for antibacterial drugs,  
 CC including anti-P. aeruginosa drugs, as templates for recombinant  
 CC production of P. aeruginosa-derived peptides or polypeptides, as target  
 CC components for diagnosis and/or treatment of P. aeruginosa-caused  
 CC infection, and in detection of P. aeruginosa sequences or other sequences  
 CC of Pseudomonas species using biochip technology. Sequences ABD01397-  
 CC ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:  
 CC The sequence data for this patent did not form part of the printed  
 CC specification but was obtained in electronic format from USPTO at  
 CC seqdata.uspto.gov/sequence.html  
 XX  
 XX Sequence 3132 BP; 496 A; 1133 C; 993 G; 510 T; 0 U; 0 Other;  
 SQ  
 Query Match 3.6%; Score 69.4; DB 11; Length 3132;  
 Best Local Similarity 43.8%; Pred. No. 0.0021;  
 Matches 454; Conservative 0; Mismatches 571; Indels 12; Gaps 3;  
 521 TGGCGGTGCGGACTTCTACGTAGCGGCAAGTACGACCTGCGCCCTCGCGCAGATGCCCT 580  
 Db 1799 TCGGCGTCGACGGCAGCAACCGACCTCAACACCGCGCGCTGCTGATCAACCTCAAGC 1858  
 Qy 581 GGTGCGCGCTTTCTTCTGGGATCGATCTTCGACATCGACACATTCGATCGGCCCCG 640  
 Db 1859 CGCAGCGAGCGCGAGCTCACCGCCAGCGAAGTGATCCAGCGCTCGAGCCCGAACTCG 1918  
 Qy 641 AGCCCGCGCTACTTCGAGCAGAAGCTGGAGCGCGTGTGGAGCAGAAACACCCAGGTT 700  
 Db 1919 ACACCTTCCCGGATCAAGCTGTATCGACGCGGTGACGACCTGACCATCGAGGACC 1978  
 Qy 701 CGATCCCTACGTCCGCTGGCTGCGCGCCCTTACTACTGTGTGATTTGGCGCGCTGCCCA 760  
 Db 1979 GGGTCGCGCGCACCCAGTACCAGTTTCACTTTCAGGAGCGCGACCGGACGTCGTGCGCG 2038

Qy	761	GCCTTGGCCAGCGCCCTGACGAGGAGCGCGTGGAGCGCGCGCCATGTTCCCGCCACCT	820
Db	2039	AGTGGGTGCCGAGAGCTGGTGGCGCGCTGCAGGAGTTGGCGAGCTCGCGAGCTGCCCA	2098
Qy	821	TCCTGTACACGAGTCGTGGGAGGACCCCGAGCCGGATATGGAGGTGATGAGATCAACC	880
Db	2099	CGGACTGGCAGGACAGGGCTTGCAGGCTTACCTGAACATCGACCGG-----ACACCG	2152
Qy	881	CMAGGACAGGTGCTGACCTGACTAGCGGGGCTGCAATGCCCTGAACTCTCTGCTGCTGC	940
Db	2153	CTTCGCGCTCGGCGTGAAGCTCTCCGACATCGACAGCGTCTTCAACAGCGCTTCGGCC	2212
Qy	941	AGGGGCGCGCGCAGGTGGTGGTGGATGCACTGCAACCCCGCGCAGTGGCGCTTCTGAGC	1000
Db	2213	AGCGGTGATCTCGACCATCTTCAACCGAGCCACCCAGTACCGGTGGTGGAGGTGG	2272
Qy	1001	TGAAGAAAGTGGCCATTTACAGAGCTGGAGTTTGAAGACGTGTGGAGCTTCTGGCGAGG	1060
Db	2273	CGCGCAGTTCCAGCTCGGCCCGCAGGCCCTCGAGCAGCTCTACGTGCGCTCCAGCGACG	2332
Qy	1061	CGGTGACCGCGCATTTGAGGAGCTGTACGAGAGAGCTGGCGCCCTTCTCTGCGCAA	1120
Db	2333	GCACCCAGGTGGCGCTGTGAGCTTGGCGAAGTGGAGAGCGGCATACCTCTGTCGCGA	2392
Qy	1121	CAAGCCACAACTTCTGGTCCAAAGCGCTCTGTGTACTTCCAGCACGGCTCTGTATCCAGG	1180
Db	2393	TCACCATATCGCCAGTTC---CCCTCGGAGACCTGTGCTTCAACTGGCCAAAGGTT	2449
Qy	1181	CGCGCATGGCAAGCTGTGGTGTCTGAGTGCCTTGGCCGTGTGGTGTGGTGTGGGCA	1240
Db	2450	ACTCCCTGGCGAGGCGGTGAGGCGATCGTGGCGCTCGAGGCGCAGCTGGAGTGGCGC	2509
Qy	1241	AGACCGTCAAGCGCTTGGCCAAAGCGCCCAACATGAGGAGCAGCGCGTCTGTGGACA	1300
Db	2510	TGAGCATGCAAGGCGAGCTTCCGCGCGCGCGCTGGCGCTTTCGAGGCGCTCGCTGTGAA	2569
Qy	1301	GCAACATGCTCATCC--ACTTGTGAAGAAACGGCGCCCAAGCGCTGGTGTGGCTGTTCG	1357
Db	2570	CGCTGTCTGTGATCTCGCTCGCTGGTGTGACCATGTACATCGTCTGGGATCTCTACG	2629
Qy	1358	TCAAGTTTGTGAGCTGTGTCTTCAACAAAGCGCGTGTGTGTGGTTCGGCGGGCGCTGC	1417
Db	2630	AGAGCTTTCATCCATCCGTGACCATCTCTCGACCTCGCTCGCGCGCGGTGGCGCGC	2689
Qy	1418	CGGCGAAGCAGTAGCGCTGATCAAGGCGCAGCGCATCCCATTTGAACTAGATCGCGC	1477
Db	2690	TGCTGGCGCTGATGTCTGGCGGGCAGGAGATCGGCATCTGTGGCGATCATCGGCATATCC	2749
Qy	1478	GCACCATGAGCGCGTGGCGGAGAACTCGCACCGTGGCGCAAGCAGAACTACTTCTACTACA	1537
Db	2750	TGCTGATCGCATCTGTCAAGAAAGACCGCATCATGATGATGATGATGATGATGATGATG	2809
Qy	1538	ACTGCTTCAACCGCAAG	1554
Db	2810	AGCGCAACGAAGGCAAG	2826

Search completed: March 13, 2006, 21:32:10  
 Job time : 1201 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: March 14, 2006, 01:29:45 ; Search time 49.8 Seconds  
(without alignments)  
5516.721 Million cell updates/sec

Title: US-10-620-914-44  
Perfect score: 3619  
Sequence: 1 acggggtcgggtcgtagcgg.....gagccagaaggaactaa 1947

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 4332886

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlh  
-O=/abes/ABSSWEB\_spool/US10620914/runat\_13032006\_101925\_5419/app\_query\_faeta\_1  
-DB=UniProt -QFMT=fastan -SUPRTX=n2p.rup -MINMATCH=0.1 -LOOCLD=0 -LOOEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=bl6umg2 -TRANS=human40.cdd -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abes05h  
-USER=US10620914 @CGN 1 1 192 @runat\_13032006\_101925\_5419 -NCPU=6 -ICPU=3  
-NO WMAP -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt 05.80.\*  
1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3441	95.1	666	2 Q6DN05	CHLRE
2	953	26.3	835	2 Q4IPW6	GIBZE
3	924	25.5	944	2 Q4PPF5	USTWA
4	910	25.1	741	2 Q6CEB3	YARLI
5	890.5	24.6	752	2 Q59TS3	CANAL
6	890.5	24.6	790	2 Q6CX18	KLJLA
7	881.5	24.4	832	2 Q52GF6	MAGGR
8	881	24.3	820	2 Q55S70	CRVNE
9	864.5	23.9	763	2 Q6BK65	DEBHA
10	850.5	23.5	908	2 Q7SGY5	NEUCR
11	845.5	23.4	797	2 Q5KX93	CRVNE
12	806.5	22.3	831	2 Q4X175	ASPFU
13	752	20.8	790	2 Q5B0L6	EMENI
14	463	12.8	415	2 Q7UYV8	RHOBA
15	413.5	11.4	767	2 Q4NR00	_9DELT
16	409.5	11.3	251	2 Q7UYV9	RHOBA

C	17	401	11.1	2268	2	Q4NRAS_9DELT	Q4nr5 anaeromyxob
	18	380.5	10.5	407	2	Q6NEV1_RHOSH	Q6nbv1 rhodopsedo
	19	379.5	10.5	416	2	Q93TQ1_RHOSH	Q93tq1 rhodobacter
	20	368	10.2	767	2	Q4NR00_9DELT	Q4nr00 anaeromyxob
	21	364	10.1	964	2	Q4NOB4_9DELT	Q4nqb4 anaeromyxob
	22	357	9.9	2147	2	Q4NUU3_9DELT	Q4nuu3 anaeromyxob
	23	354.5	9.8	1122	2	Q4NQW1_9DELT	Q4nqw1 anaeromyxob
	24	354	9.8	1286	2	Q9VR49_DROME	Q9vr49 drosophila
	25	353	9.8	1549	2	Q4NU10_9DELT	Q4nu10 anaeromyxob
	26	349	9.6	905	2	Q4NTZ7_9DELT	Q4ntz7 anaeromyxob
	27	345.5	9.5	608	2	Q9Q5K9_CHV12	Q9q5k9 cercopithec
	28	340	9.4	761	2	Q4NT26_9DELT	Q4nt26 anaeromyxob
	29	338	9.3	1549	2	Q4NU10_9DELT	Q4nu10 anaeromyxob
	30	336.5	9.3	964	2	Q4NRUS_9DELT	Q4nrus anaeromyxob
	31	335	9.3	863	2	Q4NT95_9DELT	Q4nt95 anaeromyxob
	32	333	9.2	2268	2	Q4NRAS_9DELT	Q4nr5 anaeromyxob
	33	328	9.1	964	2	Q4NOB4_9DELT	Q4nqb4 anaeromyxob
	34	327	9.0	890	2	Q8UZE1_9GAMA	Q8uz1 cercopithec
	35	326.5	9.0	416	2	Q8UDK6_AGRF5	Q8udk6 agrobacteri
	36	326	9.0	695	2	Q4NX38_9DELT	Q4nx38 anaeromyxob
	37	323.5	8.9	660	1	YHL1_EBV	P03181 Epstein-Bar
	38	323.5	8.9	660	2	Q777A3_9GAMA	Q777a3 human herpe
	39	323	8.9	953	2	Q4NMX3_9DELT	Q4nm3 anaeromyxob
	40	322.5	8.9	706	2	Q41972_MHV68	Q41972 murid herpe
	41	322.5	8.9	880	2	Q4NU98_9DELT	Q4nu98 anaeromyxob
	42	321.5	8.9	774	2	Q41971_MHV68	Q41971 murid herpe
	43	321	8.9	432	2	Q98K98_RH110	Q98k98 rhizobium 1
	44	318	8.8	2013	2	Q4NWX4_9DELT	Q4nwx4 anaeromyxob
	45	316	8.7	1309	2	Q4NU17_9DELT	Q4nu17 anaeromyxob

ALIGNMENTS

RESULT 1

Q6DN05 CHLRE  
ID Q6DN05\_CHLRE PRELIMINARY; PRT; 666 AA.  
AC Q6DN05;  
DT 25-OCT-2004 (TrEMBLrel. 28, Created)  
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)  
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
DE Betaine lipid synthase.  
GN Name=BTAL;  
OS Chlamydomonas reinhardtii.  
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;  
OC Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.  
OX NCBI\_TaxID=3055;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=CC125;  
RX PubMed=15701786; DOI=10.1128/EC.4.2.242-252.2005;  
RA Riekhof W.R., Sears B.B., Benning C.;  
RT "Annotation of Genes Involved in Glycerolipid Biosynthesis in  
RT Chlamydomonas reinhardtii: Discovery of the Betaine Lipid Synthase  
RT BTALCr.";  
RL Eukaryot. Cell 4:242-252(2005).  
DR EMBL; AY656806; AT72342.1; -, mRNA.  
DR GO; GO:0008757; F:S-adenosylmethionine-dependent methyltransf.; IEA.  
DR GO; GO:0016740; F:transferase activity; IEA.  
DR InterPro; IPR002114; Hpr SerP\_S.  
DR InterPro; IPR000051; SAM\_bind\_S.  
DR PROSITE; PS00589; PTS\_HPR\_SER; UNKNOWN\_1.  
KW Methyltransferase; Transferase.  
SQ SEQUENCE 666 AA; 75780 MW; 4B6A6F5718682284 CRC64;

Alignment Scores:  
Pred. No.: 114e-169 Length: 666  
Score: 3441.00 Matches: 647  
Percent Similarity: 97.3% Conservative: 1  
Best Local Similarity: 97.1% Mismatches: 0  
Query Match: 95.1% Indels: 18  
DB: 2 Gaps: 1

US-10-620-914-44 (1-1947) x Q6DN05\_CHLRE (1-666)

```
QY 1 ATGGGGTGGGTGGTGGAGCGCGCTGCGAGCTACACCAAGAACTTCTCCCTGGAG 60
Db 1 MeGlySerGlyArgAspGlyArgProAlaSerTyrThrIysAsnPheSerLeuGlu 20
QY 61 AAGCTCAAGCTCAGCAGCATGAAGGATGACCTGACCGTCTCTGGCCATATGTGGTGGC 120
Db 21 LysLeuLysLeuSerSerMetLysAspAspLeuThrValLeuArgHisMetTrpPheGly 40
QY 121 AGCAAGAGGGCGATGATCATCGTCTGCTGGCTGGAGAGCTTACGGGCCCCAGGCGCT 180
Db 41 SerLysLysGlyAspAspHisAlaAlaArgLeuGluSerPheTyrGlyProGlnAlaAla 60
QY 181 GCCTTT----- 186
Db 61 AlaTyrAspAlaPheArgSerArgPheLeuTrpGlyArgArgProMetLeuAlaAlaVal 80
QY 187 GCTGCCCCCTGGCCGAGCGCTCGAACCTCATCTGGGTTCACCTGGGTGGTGGCACTGGG 246
Db 81 AlaAlaArgLeuAlaGluArgSerAsnLeuIleTrpValAspLeuGlyGlyThrGly 100
QY 247 GAGAATGCTGATGATGCTGATTAATCATCGACTGGCGAAGTTCAGTCCATCTACGTG 306
Db 101 GluAsnValAspMetMetAlaAspTyrIleAspLeuAlaLysPheLysSerIleTyrVal 120
QY 307 GTCGACCTGCGCACTCGCTGTCGAGGTGCGCAAGAGGAGCGCAAGGCGCAAGGGCTGG 366
Db 121 ValAspLeuCysHisSerLeuCysGluValAlaLysLysAlaLysAlaLysGlyTrp 140
QY 367 AAGAATGCTCAGGTGCTGGAGGCGCGCTTGCCAAATTTGCGCCCTCGAGGCGCCGCG 426
Db 141 LysAsnValGlnValValGluAlaAspAlaCysGlnPheAlaProProGluGlyThrAla 160
QY 427 AGCTCATACCTTCTCTACTGCTCATCGATGATTCACCGTTCACCAAGCTCATCGAC 486
Db 161 ThrLeuIleThrPheSerTyrSerLeuThrMetIleProPheHisAsnValIleAsp 180
QY 487 CAGCTTGTCTGCTACCTGCTCCAGAGCGCTGGTGGCGTTCGCGACTTCTACGTGAGC 546
Db 181 GlnAlaCysSerTyrLeuSerGlnAspGlyLeuValGlyValAlaAspPheTyrValSer 200
QY 547 GGCAAGTACGACCTGCGCCCTGCGCAGATGCGCTGCTGGCGCGTTCCTTCGCGCATCG 606
Db 201 GlyIysTyrAspLeuProLeuArgGlnMetProTrpSerArgPhePheTrpArgSer 220
QY 607 ATCTTCGACATCGACACATTGACATCGGCCCGCGAGCGCGCGCTACCTGAGCAGAG 666
Db 221 IlePheAspIleAspAsnIleAspIleGlyProGluArgArgAlaTyrLeuGluGlnLys 240
QY 667 CTGGAGCGCGTGTGGAGCGAGACACCCAGGGTTTCGATCCCTACGTCGCTGGCTCGC 726
Db 241 LeuGluArgValTrpGluGlnAsnThrGlnGlySerIleProTyrValProTrpLeuArg 260
QY 727 GCGCCCTACTACGTGTGGATTGGCGCGCTGCGCCAGCGTTCGCCACGCGCTGCACGAGG 786
Db 261 AlaProTyrTyrValTrpIleGlyArgLeuProSerValGlyHisAlaLeuHisGluGlu 280
QY 787 CGCTGTGAGCGCGCGCCCATGTTCCCGCCACCTTCTGTCACGCGAGTGTGGGAGGAC 846
Db 281 ArgValGluArgProProMetPheProProThrPheLeuTyrThrGlnSerTrpGluAsp 300
QY 847 CCCGAGCGGATATGGAGGTGATGAGATCAACCCCAAGGACGCGTTCGACCTGACT 906
Db 301 ProGluProAspMetGluValMetGluIleAsnProLysAspThrValLeuThrLeuThr 320
QY 907 AGCGGCGGCTGCAATGCCCTGAACCTGTGTCGAGGGGCGCGCCAGAGTGTGTCGGTG 966
Db 321 SerGlyGlyCysAsnAlaLeuAsnLeuValGlnGlyAlaGlyGlnValIleValSerVal 340
QY 967 GACTGCAACCGCGCGAGTGGCGCTTCTGGAGCTGAAGAGGTGGCATTCAGCAGCTG 1026
Db 341 AspCysAsnProAlaGlnSerAlaLeuLeuGluLeuLysValAlaIleGlnGlnLeu 360
```

```
QY 1027 GAGTTTGGAGCAGCTGTGGCAGCTGTTTCGGCGAGGGCGTGCACCCGCGCATTTGAGGAGCTG 1086
Db 361 GluPheGluAspValTrpGlnLeuPheGlyGluGlyValHisProArgIleGluGluLeu 380
QY 1087 TAGCAGAGAGCTGTGGCGCTTCTCTGTCGCAAAACAGCACCAACTTCTGGTCCAGCGC 1146
Db 381 TyrGluLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrpSerLysArg 400
QY 1147 CTCTGCTACTTCCAGCAGCGCTGTACTACAGGGCGGATGGCAAGCTGTGCTGGGTG 1206
Db 401 LeuTrpTyrPheGlnHisGlyLeuTyrTyrGlnGlyLysLeuLysCysTrpVal 420
QY 1207 CTCAGTGTCTGCGCGCTGTGGGACTGGGCAAGACCGTCAAGCGCTTCGCGCAACGCG 1266
Db 421 LeuGlnCysLeuAlaValValLeuGlyLeuGlyLysThrValLysArgLeuAlaAsnAla 440
QY 1267 CCACAATGAGAGGAGCAGCGCTCTGTGTGGGACAGCAACATGCTCATCTCCTGCTGAAG 1326
Db 441 ProThrMetGluGluGlnArgArgLeuTrpAspSerAsnMetLeuIleHisPheValLys 460
QY 1327 AACGGCCCAAGCGCTGTGTGGTGTCTCAAGTTCTGTGAGCTGTGTCTCTCAAC 1386
Db 461 AsnGlyProLysProLeuValTrpLeuPheValLysPheValSerLeuValLeuPheAsn 480
QY 1387 AAGCGCTGTGTGTGGTTCGGCGCGCTGCGCGCAAGCAGTACGCGCTGATCAAGCGC 1446
Db 481 LysAlaValLeuTrpPheGlyGlyGlyValProGlyLysGlnTyrAlaLeuIleLysAla 500
QY 1447 GAGCGATCCCATTTGAGAACTATACGCGCGCGCACCATGAGCGCGTGGCGGAGAACTCG 1506
Db 501 AspGlyIleProIleGluAsnTyrIleAlaArgThrMetAspGlyValAlaGluAsnSer 520
QY 1507 CAGTGTGGCAAGCAGAACTACTTCTACTCAACTGCTCACCAGTTCCTGCGCGGAC 1566
Db 521 HisValArgLysGlnAsnTyrPheTyrTyrAsnCysLeuThrGlyLysPheLeuArgAsp 540
QY 1567 AACTGCCCCACCTACCTGCGCGAGCGCTTCGCCACCTCAAGAGTGGCGTGGTGGAC 1626
Db 541 AsnCysProThrTyrLeuArgGluAlaAlaPheAlaThrLeuLysSerGlyValValAsp 560
QY 1627 AACTGACCGCTCCACCAACTTCTCATGAGAGCTCAAGGGCGCCTACACCAAG 1686
Db 561 AsnLeuThrValSerThrAsnPhePheMetGluGluLeuLysAlaArgThrTyrThrLys 580
QY 1687 GTGATTTCTGATGACCAACAGCTGGGCTGCTGATATGCCCTGGCCACAGAGTGGCGCAG 1746
Db 581 ValIleLeuMetAspHisValAspTrpLeuAspMetProValAlaAsnGluLeuAlaGlu 600
QY 1747 TGCTTGCCCAAGCAGGTGTGCGCGCGGCGCATCTCATCTGGCGCTCCGCTCCCTCAGC 1806
Db 601 CysLeuAlaLysGlnValAlaProGlyGlyIleValIleTrpArgSerAlaSerLeuSer 620
QY 1807 CCGCCTTACCGCGAGCTGATCCAGAGCGCGCTTGCAGCTGGCTGCATCCGCGCGGCC 1866
Db 621 ProProTyrAlaGluLeuIleGlnLysAlaGlyPheAspValArgCysIleArgArgAla 640
QY 1867 ACTCAGGCTTACATGGACCGCTCAACATGTACAGCTCTTCTACATGGCGCGCGCGAAG 1926
Db 641 ThrGlnGlyTyrMetAspArgValAsnMetTyrSerSerPheTyrMetAlaArgLys 660
QY 1927 GCGCCCAAGAGGACCAAC 1944
Db 661 GlyAlaLysLysAspAsn 666
RESULT 2
Q4IPW6 GIBZE
ID Q4IPW6 GIBZE PRELIMINARY; PRT; 835 AA.
AC Q4IPW6;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Hypothetical protein.
GN ORFNames=FG00742.1;
```



Db	183	PheArgLeuGluAspTyrGluSerGlyMetProSerLysSerIleProProArgSerPro	202
Qy	418	-----GGCACCCGCGAGCTCATCACCTTC	441
Db	203	AlaLeuSerTyrPheAspLysProArgProAspPheGlyAlaAspLeuIleThrMet	222
Qy	442	TCTACTCGCTCAGCATGATCCACCGTTCACACAGTCATCAGCACGAGGCTTCTCGTAC	501
Db	223	SerTyrSerLeuSerMetIleProAspTyrTyrSerValIleAspSerValThrSerLeu	242
Qy	502	CTCTCCCAAGACGGCTGGTGGCGCTTCCGACTCTACGTGAGCGGACGACGACCTG	561
Db	243	LeuSerProGlnGlyIleMetGlyValValAspPheTyrValGlnAenLysValAspPhe	262
Qy	562	CCCTCG-----CGCCAGATGCCCTGGTGCCTGCCCGCTTC	594
Db	263	AlaPheArgAsnTyrThrGlyGlyLeuValAspArgHisValAsnPheLeuSerArgSer	282
Qy	595	TTCTGGCGATCGCATCTTCGACATCGACAATGACATCGCCCGGAGCGCCGCGCTAC	654
Db	283	PheTrpArgSerTrpPheAspLeuAspArgValGlyLeuGluProSerArgArgAspTyr	302
Qy	655	CTGGACGAGAGCTGGAGCGGTGTGGAGCAGACAACCCAG-----GGTTCG	702
Db	303	LeuGluTyrLysPheGlyThrValLeuAsnValAsnThrArgAsnLysGlyLeuGlyAla	322
Qy	703	ATCCCCCTACGTGCGTGGCTGCGCGCCCTACTAGTGTGGATTGCG-----CGCCTG	756
Db	323	Ile-----ProTyrTyrIleTrpLeuGlyCysHisLysLys	334
Qy	757	CCCAGCGTTGGCCACGCCCTGCAGGAGCGGTGGAGCGG-----	798
Db	335	ProPheSerSerSerLeuProHisGluIleValGluArgIleAspAlaLeuValThr	354
Qy	799	-----CCGCCCATGTTCCCG-----	813
Db	355	GluSerProTyrLeuTyrProAlaAsnHisGlyAspAlaLeuThrArgAlaIleGluArg	374
Qy	813	-----	813
Db	375	SerAlaProGluIleArgSerLysAlaPheLeuThrAlaValSerAsnLeuSerSerAsn	394
Qy	814	-----CCCACTTCCTGTAC-----	828
Db	395	LeuProLeuProSerPhePheTyrGlnAsnHisHisTrpArgIleTyrTyrAspGluGln	414
Qy	829	-----ACGACG-----TCGTGGGAGGACCCC	849
Db	415	LeuProLysHisThrGlnPheLysAspGluTyrIleTyrAlaPheThrTrpGluAspThr	434
Qy	850	GAGCCGGATATGGAGGTGATGGAGATCAACCCCAAGCAGACGGTGTGTGACCTGACCTAGC	909
Db	435	ArgValAspGluArgIleLeuLysLeuGlyAlaAspAspLysValLeuAlaIleThrSer	454
Qy	910	GGCGGTGCAATGCCCTGCACTGTGTGTGTCAGGGGGCGGCGCAGCTGTGTGTCGTGGAC	969
Db	455	AlaGlyAspAsnIleLeuSerTyrLeuLeuGlnSerProAlaArgValHisAlaValAsp	474
Qy	970	TGCAACCCCGCAGTGGCGCTTCTGGAGCTGAAGAAGGTGGCCATTACGACGTGGAG	1029
Db	475	LeuAsnProThrGlnAenHisLeuLeuGluLeuLysAlaAlaSerTyrThrAlaLeuPro	494
Qy	1030	TTTGAGGAGCTGGGAGCTGTTCGGCGAGGGCGTGCACCCCGGCTATGGAGGCTGTAC	1089
Db	495	TyrGlyAspPheTrpLysIlePheGlyAspGlyLysHisProArgPheArgGlyLeuLeu	514
Qy	1090	GAGAAAGAGCTGGCGCTTCTGTGTGCGCAACCCAGCACAACCTCTGGTCCAAAGCGCTC	1149
Db	515	IleThrLysLeuSerProHisLeuSerGlyArgAlaPheGlnTyrTrpLeuLysAsnVal	534
Qy	1150	TGTTACTTTCAG-----CACGGCTCTACTACGAGGCGGCGATGGGCAAGCTG	1197



Db 239 GlyGlyThrGlyTrpAsnIleGluGlnMetAsnGlnPhePheProIleAspGlnLeuSer 258  
Qy 295 TCATCTACGTGTCGACGTGTCACCTCGCTGCGAGGTGGCGCAAGAAAGCGCAAG 354  
Db 259 GlnValTyrLeuIleAspLeuGlnValProLeuGlnValAlaArgLysArgPheAla 278  
Qy 355 GCCAAGGCTGGAGATGTCAGGTGTCGAGCGCCGACCTTGCACATTTGGCCCT 414  
Db 279 AlaLysGlyPheLysAsnValGlnValLeuCysGlnAspAlaSerGlnPheAsnMetPro 298  
Qy 415 GAGGCGACCGCGAGC-----CTCATCACCTTCTCTCTCTCTCTCTCTCTCTCT 459  
Db 299 ---GlyLeuAlaAlaGlyGlnLysValAspLeuPheThrCysSerTyrSerIleSerMet 317  
Qy 460 ATTCCACCGTCCACACGTCATCGACCGCTTGTCTGCTACCTGTCCCAA---GACGGC 516  
Db 318 IleProPheTyrAlaValLeuAspArgAlaAsnAspLeuLeuAspProValThrGly 337  
Qy 517 CTGGTGGCGCTTGGCGACTTCTAGCTGAGCGGCAAGTACACCTGCCCTG----- 567  
Db 338 ValPheGlyValValAspPheTyrValSerGlySerSerGly---ProLeuAlaLysSer 356  
Qy 568 -----CGCCAGATGCCCTGCTGCGCGCGCTTCTCTCTCTCTCTCTCTCTCT 603  
Db 357 ProMetIleGlyGlyAspThrArgArgGlnCysGlyTyrLeuSerArgTrpPheTrpSer 376  
Qy 604 TCGATCTTCGACATCGACACATTCATCGCCCGCGCGCGCTACCTGAGAGCAG 663  
Db 377 MetTrpPheSerPheAspHieGluLeuHisProAlaArgAspTyrLeuGluHis 396  
Qy 664 AAGCTGGAGCGGTGGGAGCAGACACACCGAGGTTCG-----ATCCCTACGTGCGG 717  
Db 397 LysPheGlyThrIleLysCysTyrAsnGlyArgAsnAsnPheIleIleProPheIle--- 415  
Qy 718 TGCTGGCGCCCTTACTACGTGTGATTGGC-----CGCTGCCCGCGTGGCCAC 771  
Db 416 ---ValArgIleProTyrIleIleTrpLeuGlyValSerArgGluArgPheThrLys 434  
Qy 772 GCCTGCGAGGAGCGCGTGGAGCGG-----CGCCCATGTTCCCG 813  
Db 435 AlaIleGlnAlaPheGluValGluSerGlyAsnArgValValProProSerPhePro 454  
Qy 814 CCC----- 816  
Db 455 GluLeuAlaTyrMetHisAsnGlyGluAlaHisThrHisSerIleSerAlaAlaThr 474  
Qy 816 ----- 816  
Db 475 AlaThrAlaThrAlaThrGlyIleAspAlaAspGluAspAlaSerSerArgPhe 494  
Qy 816 ----- 816  
Db 495 GluProAlaArgArgLeuLeuArgArgValSerGluAlaThrThrGluSerAspAla 514  
Qy 816 ----- 816  
Db 515 AspSerAspArgProLeuLysLeuGluLeuGlyProHisPheProLeuSerSerPheHis 534  
Qy 816 ----- 816  
Db 535 TyrGlnLysArgGlnTrpArgLeuProPheValAspAsnGluPheSerAspMetPheArg 554  
Qy 817 ACCTTCTGTACACGAGTCTGGGAGGACCCGAGCGCGATATGGAGTGTGAGATC 876  
Db 555 ThrTrpIleTyrGlyPheThrTrpGluAspProTyrValAspMetGlnHisLeuAspLeu 574  
Qy 877 AACCCCAAGACAGCGTGTGACCTTCAGTACGCGCGGTGCAATGCCCTGAACCTGCTG 936  
Db 575 GlyLysAspAspSerIleLeuCysIleThrSerAlaGlyAspAsnAlaLeuHisTyrAla 594  
Qy 937 GTCACAGGG---CGCCGCCAGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 993  
Db 595 ValAlaGlyLysProArgArgIleHisAlaValAspMetAsnProCysGlnGlyHisLeu 614

RESULT 4

Q6C6E3 YARLI

ID Q6C6E3\_YARLI PRELIMINARY; PRT; 741 AA.

AC Q6C6E3;

Qy 994 CTGAGCTGAAGAAGTGGCCATTTCAGCAGCTGGAGTTTGAGGACGTCGTGCGAGCTGTC 1053  
Db 615 LeuGluLeuLysLeuAlaCysIleAlaSerLeuSerTyrAspGluMetTrpGlnMetPhe 634  
Qy 1054 GCGAGGCGCTGCACCCGCCATTGAGGAGCTGTACGAGAGAGCTGGCGCGCTTCTCTG 1113  
Db 635 GlyGluGlyArgIleAspAsnPheArgGluLeuAspSerLysIleSerProTyrLeu 654  
Qy 1114 TCCAAACAGCAGCCACCACTTCTGCTCCAGCGCTCTGCTACTTCCAGCAGCGCTCTAC 1173  
Db 655 SerSerHisAlaTyrGlnPheTrpArgLeuAsnThrArgAlaPheAspLysAlaPheTyr 674  
Qy 1174 TACAGGCGGCGATGGGCAAGCTGTGCTGGGTGCTGCAGTGCCTGCGCGTGTCTGGGA 1233  
Db 675 PheArgGlyTyrSerGlyHisAlaLeuArgLeuAlaLysPheAlaPheSerValThrGly 694  
Qy 1234 CTGGGCAAGACCTCAAGCGCTCCGCAACGCGCCCAATGAGGAGGAGCGCGCTG 1293  
Db 695 ValArgArgTrpValGluLysMetCysThrAlaAsnSerValGluGluGlnGlnVal 714  
Qy 1294 TGGGACAGCAACATGCTCATCCACTTCTGTAAGAGCGGCGCCCAAGCGCTGGTGGCTG 1353  
Db 715 TrpAspLysLysLeuArgSerThrLeuIleAsn-----LysProLeuIleArgLeu 731  
Qy 1354 TTCGTCAAGTTCGTGACCTGTGCTTCAACAAGCGCTGCTGTGGTTCGGCGCGCG 1413  
Db 732 Phe-----LeuSerAsnProAlaPheLeuTrpAsnAlaLeuGly 744  
Qy 1414 GTGCGCGCAAGCAGTACCGCTGATCAAGCGGAGGCGATCCCATTTGAGACTACATC 1473  
Db 745 ValProMetAsnGlnTyrGlnIlePheLeuAsnGluGlyValSerAlaGluGlnPheAla 764  
Qy 1474 GCGCGCACATGACGCGCTGCGGAGAACTCCAGCTGCGCAAGCAGACTACTTCTAC 1533  
Db 765 IleAspThrLeuAspSerIleProSerArgSerLeuIleLysAsnAspAsnTyrHisTyr 784  
Qy 1534 TACAATGCTCACCGGCAAGTTCCTGCGGACAACTGCCCCACCTACCTGCGCGAGCG 1593  
Db 785 GlnLeuCysLeuLeuHisLysTyrThrLysGlnSerCysProLeuTyrLeuLysProAsp 804  
Qy 1594 GCCTTCCGACCTCAAGAGTGGCGTG-----GTGACCAACTGACCGTCTCC 1641  
Db 805 GlyPheAlaLeuLysLysGlnAlaLeuGlnAspGlyLeuAspSerPheArgLeuHis 824  
Qy 1642 ACCAATCTTCTCATGAGGAGCTCAAGCG-----CGCACCTACACCAAGTATT 1692  
Db 825 ThrAspSerIleValAsnValLeuArgGlyPheGluAspGlyAlaLeuThrArgAlaIle 844  
Qy 1693 CTGATGACGACCTGACCTGGCTGGAT---ATGCCCCGTGGCC----- 1731  
Db 845 ThrMetAspHisMetAspTrpPheAspProValProAlaSerArgProAlaProThrIle 864  
Qy 1732 -----AACGAGCTGGCGAG 1746  
Db 865 LysGlnAlaArgAspSerAspLysSerValSerAspLeuAspArgGluIleCysGlu 884  
Qy 1747 TGCTGGCGCAAGCAGGTTCGCGCGCGCATCTGCGGCTGCGGCTGCGGCTGCGGCTGCG 1806  
Db 885 ---LeuSerArgValIleArgLysGlyAlaValPheTyrArgSerAlaAlaLysLys 903  
Qy 1807 CGCCCTACCGCGAGCTGATCCAGAGCGGCTTCGAGTGCCTGCGATCCGC---CGC 1863  
Db 904 ProTrpTyrAsnHisArgPheGluLysMetGlyPheSerValGlnProValHisIleArg 923  
Qy 1864 GCCACTCAGGCTACATGAGCGCGTCAACATGATACAGTCTCTTACATGGCGCGCGCG 1923  
Db 924 GluThrAlaLysProIleAspAsnValAsnMetTyrAlaSerPheTyrLysAlaThrArg 943  
RESULT 4  
Q6C6E3 YARLI  
ID Q6C6E3\_YARLI PRELIMINARY; PRT; 741 AA.  
AC Q6C6E3;





```

QY 819 ----- 819
Db 322 ThrProIleAlaAsnGlnLeuGluAspIleProIleSerLysGlyHisGluAlaAlaLeu 341
QY 819 ----- 819
Db 342 IleAsnLeuGlnLysAsnLeuProTyrProSerMetTyrTyrGlnLysGluTyrTrpArg 361
QY 820 ----- 820
Db 362 ValTyrTyrAspGluMetAsnProLeuTyrGluGlnPheLysAsnGlnTyrIleTyrAla 381
QY 832 CAGTCGGGAGGAGCACCAGCGGATATGAGGTGATGAGGTGATGAGATCAACCCCAAGGACACG 891
Db 382 PheThrTrpGluAspProArgGluAspHisLysLeuAsnPheThrSerAspAspThr 401
QY 892 GTGCTGACCTGATAGCGCGGTGCAATGCCCTGAACCTG-----CTGGTGACGGG 945
Db 402 ValLeuAlaIleThrSerAlaGlyAspAsnIleLeuSerTyrAlaSerLeuProThrPro 421
QY 946 GCCGGCCAGTGTGTGCTGGTGAAGTCAACCCCGCAGTCGGCGCTCTCGAGCTGAAG 1005
Db 422 ProLysLysIleHisAlaValAspLeuAsnProCysGlnAsnHisLeuGluLeuLys 441
QY 1006 AAGGTGGCCATTACAGCAGCTGGAGTTTGAGGAGCTGTGGCAGCTGTTCGGCGAGGGCGTG 1065
Db 442 LeuAlaSerPheArgCysLeuSerGlnGluGlnIleTrpSerMetPheGlyGluGlyLys 461
QY 1066 CACCCGCGCATGTAGGAGCTGTACGAGAAGAGTGGCGCTTCCTGTCTGCCAACACGAG 1125
Db 462 IleGluAsnPheAsnAspLeuLeuIleAspThrLeuAlaProHisMetSerSerAsnAla 481
QY 1126 CACACTTCTGGTCCAGCGC-----CTCTGTGTACTTCACGACGCGCTGTACTACGAG 1179
Db 482 PheGlnTyrTrpMetAspLysGlyProLysThrPheSerGlyLysGlyLeuTyr----- 499
QY 1180 GCGGCGATGGCAAGCTGTGCTGGTGTCTGCGAGTGC-----CTGGCGGTGGTCTG 1230
Db 500 ---AspThrGlyPheSerArgTrpAlaLeuArgLeuSerArgTyrValPheLysValCys 518
QY 1231 GGACTGGGCAAGCCGTCAAGCGCTGCGCCACCGCCCAATGGAGGACGACGCGCT 1290
Db 519 GlyValSerLysTyrValGluGluLeuCysAlaAlaThrThrMetGluGluGlnLeuArg 538
QY 1291 CTGTGGACACCAATCTCTCATCTCCATCTGCTGAAGACGGCGCCAGCGCTGTGTGG 1350
Db 539 IleTrpAsnGluHisLeu-----LysProThr----- 547
QY 1351 CTGTGTCTCAAGTTCTGTG---AGCTGTGTCTCTTCAACAGGCGGTGTGTGTCTGCGC 1407
Db 548 LeuPheAsnProValValGlySerLeuLeuValGlyAsnProMetPheLeuTrpLysAla 567
QY 1408 GCGGCGGTGGCGGCAAGCAGTACGCGCTGATCAAGCGGACGACGATCCCAATTGAGAAC 1467
Db 568 LeuGlyValProAlaAsnGlnAlaLeuMet-----GlyProSerValIleLys 584
QY 1468 TACATCGCGGCCACATGAGCGGTGGCGGAGAACTGCGACGTGCGGACGACCAACTAC 1527
Db 585 TyrValValAspThrLeuAspProIleIleLysArgSerMetIleSerAsnAspAsnTyr 604
QY 1528 TTCTACTACAACTGCCTCACCGCAAGTTCTGCGCGCACTACCTCCCACTACTGCGC 1587
Db 605 PheTyrTyrLeuCysMetGlyArgTyrThrLysAsnAsnCysProAspTyrLeuThr 624
QY 1588 GAGCGCGCTTGGCCACCTCCAAAG-----AGTGGCGGTGGT 1623
Db 625 ThrLysGlyPheAsnArgLeuSerSerThrAlaAlaThrAlaSerGlySerSerProIle 644
QY 1624 GACAACTGACGCTCTCCACCAACTTCTTCATGAG-----GAGCTCAAGCGCGC 1674
Db 645 AspAsnLeuArgIleHisThrAspThrLeuAsnGluValPheGlyArgLeuLysGluLys 664
QY 1675 ACCTACCAAGGTGATTCTGATGACCACTGGAGTGGTGGAT-----ATGCC 1725

```

```

Db 665 SerIleThrIleAlaIleIleMetAspHisMetAspTrpPheAspProAsnGlyArgAsp 684
QY 1726 GTGGCCCAACGAGTGGCCGAGTGCCTGCGCCAGCAGGTTGCGCGCGGCATCGTCATC 1785
Db 685 AlaIleAsnGluIleThr---AlaLeuLysArgCysLeuAlaProGlyGlyArgValLeu 703
QY 1786 TGGCGGTCTCGCCTCCCTCAGCCCGCTACGCCGAGCTGATCCAGAGGCGGCGCTTC--- 1842
Db 704 LeuArgSerAlaSerThrLysProTyrTyrLeuLysThrPheLysAsnLeuGlyPheGln 723
QY 1843 GAGCTGCGCTGCATCCGCGCGCCATCTCAGGGCTACATGACCGCGTCAACATGTACAGC 1902
Db 724 GluGluGluAsnValValArgGlnProGlySerSerIleAspArgValAsnMetTyrAla 743
QY 1903 TCC 1905
Db 744 Asn 744

RESULT 6
Q6CJ18_KJLJLA
ID Q6CJ18_KJLJLA PRELIMINARY; PRT; 790 AA.
AC Q6CJ18;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DE Similar to ca|C1A1783|pf7635 Candida albicans unknown function.
GN OrderedLocusNames=K11A0F22198g;
OS Kluyveromyces lactis (yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX NCBI_TaxID=28985;
RN [1]
NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA]
RP STRAIN=CBS 2359 / IFO 1267 / NRRL Y-1140 / WM37;
RX PubMed=15229592; DOI=10.1038/nature02579;
RA Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,
RA Goffard N., Franchin S., Beckerich J.-M., Beyne E., Bleykasten C.,
RA Barnay S., Blanchin S., Catolico L., Confanieri F., de Daruvar A.,
RA Boisrame A., Boyer J., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Despons L., Fabre E., Fairhead C., Jaumaux N., Joyet P., Kachouri R.,
RA Hantraye F., Hennequin C., Lemaire M., Lesur I., Ma L., Muller H.,
RA Kerret A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.-M., Nikolski M., Ozas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Petier S., Richard G.-F., Straub M.-L., Suleau A.,
RA Swennen D., Tekala F., Wesolowski-Jouvet M., Westhof E., Wirth B.,
RA Zeniou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gallardin C., Weissenbach J.,
RA Wincker P., Souciet J.-L.;
RT "Genome evolution in yeasts.";
RL Nature 430:35-44(2004)
DR EMBL; CR382126; CAG98779.1; -; Genomic DNA.
DR GO; GO:0008757; F:S-adenosylmethionine-dependent methyltransf. .; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000051; SAM bind.
KW Complete proteome; Methyltransferase; Transferase.
SQ SEQUENCE 790 AA; 90212 MW; 25BA985E9C5B0D1 CRC64;

Alignment Scores:
Pred. No.: 8,14e-38 Length: 790
Score: 890.50 Matches: 233
Percent Similarity: 46.6% Conservative: 94
Best Local Similarity: 33.2% Mismatches: 241
Query Match: 24.6% Indels: 133
Gaps: 2

US-10-620-914-44 (1-1947) x Q6CJ18_KJLJLA (1-790)
QY 151 CTGAGAGAGCTTCTACGGCGCCCGCGCTGCTTT----- 186
Db 93 LeuGluGlnPheTyrLysSerGlnAlaLysLeuTyrAspArgThrArgGlyValLeuLeu 112

```



187 QY -----GCTGCCCGCTGGCGGAGCGCTGCAACCTC 216  
113 Db GlnGlyArgGluThrSerLeuLeuLeuSerHisLeuSerGluLeuLeuGlyAsn 132  
217 QY ATCTGGGTGACCTGGGTGGGACCTGGGAGAAATGTCGATATGATGGCTGATTACATC 276  
133 Db ValTrpIleAspValGlyGlyThrGlyPheAsnIleSerGlnMetAlaLeuLeuThr 152  
277 QY GACCTG---GCGAAGTTCAAGTCCATCTACGTGGTCCAGCTGTGCCACTCGCTGTGCGAG 333  
153 Db AsnLeuAspThrThrPheAspLysIleLeuLeuAspLeuSerProSerLeuLeuGlu 172  
334 QY GTGGCCAGAGAGGCGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAG 393  
173 Db ValAlaArgLysArgCysGlyGluHisGlyTrpLysAsnValGluValIleCysGlyAsp 192  
394 QY GCTTGCCAAATTGGCCCTCGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGG 453  
193 Db AlaCysAspPheGluIleProGluLeuSerAlaGlnLeuIleThrPheSerTyrSerLeu 212  
454 QY ACGATGATTCACCGTTCCACAACGTCATCGACGAGCTTGTCTGTACCTG---TCCCAA 510  
213 Db SerMetIleProSerPheTyrAlaAlaIleAspHisAlaValSerLeuLeuAspAlaLys 232  
511 QY GACGCTGTGGCGCTGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGG 555  
233 Db AsnGlyIleIleSerCysValAspPheGlyValThrAsnGluSerMetLeuValGlyArg 252  
556 QY -----GACCTGCCCTCGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGG 600  
253 Db ThrAsnThrLeuGlyGlyLeuValAsnArgHisIleProTrpLeuPheArgThrPheTrp 272  
601 QY CGATCGATTCGACATCGACAACATGACATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660  
273 Db ArgLeuTrpPheGluPheAspLysValPheLeuAspProAlaArgGluTyrLeuGlu 292  
661 QY CAGAGCTGAGGCGCTGGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGG 708  
293 Db TyrArgPheGlyThrIleLysSerLeuAsnCysTyrAsnTyrLysLeuGlyLysIlePro 312  
709 QY TACGTGCGGTGGGTG--- 723  
313 Db TyrTrpIleTrpLeuGlyCysAsnLysAspHisGluGlnHisLeuGlnAlaArgPheVal 332  
724 QY -----CGCGCCCTACTACGTGGATTGGCGG--- 753  
333 Db GluLeuAlaProThrSerProTyrLeuAlaProIleThrThrSerAlaSerSerAsnAla 352  
754 QY -----CTG 756  
353 Db GlnProMetThrLysAlaMetIleAlaAlaLeuGluAsnSerLysLysGlyLeuProTyr 372  
757 QY CCACGCTGGCCAGCGCCCTGACGAGGAGCGGTG-----GAGCGCGCGCC--- 804  
373 Db ProSerLeuPheTyrGlnLysGluHisTrpArgValTyrTrpAspGluValAsnProGlu 392  
805 QY -----ATGTTCCCGCCACTTCTGTATACGCGAGTCGTGGGAGGAGCCCGCGCGAT 858  
393 Db TyrSerGlnPheLysAsnSerTyrIleTyrAlaPheThrTrpGluAspProArgGluAsp 412  
859 QY ATGAGGTGATGAGATCAACCCAGGACACGCTGTGACCTGACCTGACGCGCGGCTGC 918  
413 Db ValAsnIleLeuAsnIleGlnProGluAspThrIleLeuAlaIleThrSerAlaGlyAsp 432  
919 QY AATGCGCTGAAC-----CTGCTGGTGCAGGGGCGGCGGAGGCGGAGGCGGAGGCGGAG 972  
433 Db AsnIleLeuHisTyrAlaThrLeuProAsnProProLysArgIleHisGlyValAspLeu 452  
973 QY AACCCCGCGGAGCGGCTTCTGGAGCTGAAGAAGGTGGCCATTTCAGCAGCTGGAGTTT 1032  
453 Db AsnProCysGlnGlyHisLeuThrGluLeuLysLeuAlaAlaIleArgSerLeuSerPhe 472  
1033 QY GAGGACGTGTGGCAGCTGTTCGGGAGGCGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAG 1092

## RESULT 7

Q52GF6 MAGGR

ID Q52GF6\_MAGGR PRELIMINARY; PRT; 832 AA.

AC Q52GF6\_

DT 13-SEP-2005 (T-EMBLrel. 31, Created)

DT 13-SEP-2005 (T-EMBLrel. 31, Last sequence update)

473 Db ThrGlnLeuTrpGlnMetPheGlyGluGlyLysIleAspArgPheAsnAsnIleLeuLeu 492  
1093 QY AAGAAGCTGGCGGCTTCTCTGTCTGCAAAACCCACCAACCTTCTGGTCCAAGCGCTCTGG 1152  
493 Db AsnLysLeuAlaProTyrLeuSerSerAsnAlaPheGlnTyrTrp----- 507  
1153 QY TACTTCCAGCACGCGCTG-----TACTACAGGGCGGC 1185  
508 Db ---PheGluAsnGlyThrLysThrPheAspProAsnGlyAlaGlyLeuTyrAspThrGly 526  
1186 QY ATGGGCAAGCTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1236  
527 Db PheThrLys-----TrpAlaLeuArgLeuAlaLysTrpValPheLysValAlaAsnLeu 544  
1237 QY GGCAAGACCGCTCAAGCGCTCGCCCAAGCCCAATGAGGAGGAGGAGGAGGAGGAGGAGGAG 1296  
545 Db ThrAspGluValAsnMetLeuCysLysAlaLysThrLeuGluGluGlnArgSerIleTrp 564  
1297 QY GACAGCAACATGCTCATCTCCACTTCTGTAAGAACGGGCGGCGGCGGCGGCGGCGGCGG 1356  
565 Db AspLysLysIle-----LysProVal-----LeuPhe 573  
1357 QY GTCAAGTTCGTGAGCTGTGTCTTTC---AACAGGCGGCTGTGTGTGTGTGTGTGTGTGTGT 1413  
574 Db AsnArgValValGlyLysIleLeuValGlyAsnProLeuPheLeuTrpSerAlaLeuGly 593  
1414 QY GTCCCGGCGAAGCAGTACGCGCTGATCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1473  
594 Db ValProArgAsnGln-----AlaLysMetMetGlySerSerThrLeuGlnTyrIle 610  
1474 QY GCGCGCACCATGAGCGGCGGCGGAGAACTCCGACGTGGCGGAGCAAGCAAGCAACTTCTTAC 1533  
611 Db IleAspThrLeuAspProValIleAspAsnSerLeuIleSerAspAspAsnTyrPheTyr 630  
1534 QY TACAACTGCTCACCAGGCTTCTGCGGACAACTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1593  
631 Db TyrLeuThrLeuLysGlyArgTyrSerSerArgSerCysProAspTyrLeuLysGluGly 650  
1594 QY GCCTTCGCGCACCTCAAG-----AGTGGGTGTGTGACCAACCTGACCGCTCTCC 1641  
651 Db GlyPheLysSerLeuSerArgGluSerProGluSerProLeuAspArgValArgLeuHis 670  
1642 QY ACCAACTTCTTC-----ATGAGGAGCTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1692  
671 Db ThrAspThrLeuLysAspValCysGluArgLysSerLysLysThrValSerIleAlaIle 690  
1693 QY CTGATGACACCGCTGAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1746  
691 Db IleMetAspHisMetAspTrpPheAsp---ProGlnGlyThrAspValAspGluIle 709  
1747 QY ---TGCTGCGCAAGCAGGTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1803  
710 Db GlnAlaLeuTrpLeuAlaLeuAsnSerArgGlyArgValLeuLeuArgSerAlaSerLys 729  
1804 QY AGCCCGCTTACCGGAGCTGATCCAGAAAGCGGCGGCTTCGACGTGGCTGCATCCGCGCGC 1863  
730 Db SerProTrpTyrIleLysAsnPheGluLysPheSerCysLysAlaValSerAla 749  
1864 QY GCCACTCAGGCG---TACATGGACCGGCTCAACATGTACAGCTCTTCTACATGGCGCGC 1920  
750 Db ArgTyrProGlyLysCysIleAspArgValAsnMetTyrAlaSerThrTrpValCysGln 769  
1921 QY CGG 1923  
770 Db Lys 770









```

Db 596 ILeSerGluAspAenTyrPheTyrTyrLeuThrLeuLysGlyLysTyrThrProGlnAen 615
    ::: ::: ||||||| ||| |||||:::
QY 1570 TGCCCACTACCTGCGCGAGCGGCTTCGCCACCTC-----1608
    ||||| |||||::: |||||
Db 616 CysProAspTyrLeuThrGluLysGlyTyrLysSerLeuThrThrLeuAenArgLysSer 635
    ||||| |||||::: |||||
QY 1609 AAGAGTGGCGTGTGGACACCTGACCGTCTCCACC-----AATCTTCTTCATGGAG 1659
    ||||| |||||::: |||||
Db 636 LysGluAlaProIleAspAenIleArgLeuHisThrAspMetLeuAenAspValPheAla 655
    ||||| |||||::: |||||
QY 1660 GAGCTCAAGCGCGCTACACAGAGTGATTCGTGATGACCACTGGCTGGCTGGAT 1719
    ||||| |||||::: |||||
Db 656 ArgLeuSerLysSerLeuSerIleAlaIleMetAspHisMetAspTyrPheAsp 675
    ||||| |||||::: |||||
QY 1720 ATGCCGTGGCCAAAGAGCTGGCGGAGTGGCTGGCCAGCAG-----GTTGGCGCGGC 1773
    ||||| |||||::: |||||
Db 676 ProAsnGlyGluAspAlaLeuAenGluIleThrAlaLeuLysSerCysLeuAenThrAen 695
    ||||| |||||::: |||||
QY 1774 GGCATCGTCACTGGCGCTCGCTCCCTCAGCGCCCTACGCCGAGCTGATCCAGAAG 1833
    ||||| |||||::: |||||
Db 696 GlyArgValLeuLeuArgSerAlaSerThrAsnProTyrPyrIleLysThrPheGluAsp 715
    ||||| |||||::: |||||
QY 1834 GCGGCTTCACGCTGCGCTGCATCCGC---CGCGCCACTCAGGCTACATGACCGCGCT 1890
    ||||| |||||::: |||||
Db 716 LeuGlyPheThrCysLysAlaAlaGlyIleArgAspThrGlyIleSerIleAspArgIle 735
    ||||| |||||::: |||||
QY 1891 AACATGTACAGCTCTCTACATGGCCCGCGG 1923
    ||||| |||||::: |||||
Db 736 AsnMetTyrAlaSerThrTyrValCysThrLys 746
    ||||| |||||::: |||||

RESULT 10
Q7SGY5_NEUCR PRELIMINARY; PRT; 908 AA.
AC Q7SGY5;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Hypothetical protein (Related to S-adenosylmethionine:diacylglycerol
DE 3-amino-3-carboxypropyl transferase btaA).
GN Name=NCU03032.1; Synonyms=B19P24.050;
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OR74A;
RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
RA Jaffe D., FitzHugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,
RA Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
RA Qui D., Ianakiev P., Pedersen D., Nelson M., Washburne M.,
RA Selitrenikoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schulte U.,
RA Kothe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D.,
RA Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gnerre S.,
RA Kanal M., Kanysseles M., Mauceli E., Bielke C., Rudd S., Frishman D.,
RA Krystofova S., Rasmussen C., Metzenberg R.L., Perkins D.D., Kroken S.,
RA Cogoni C., Macino G., Catchside D., Li W., Pratt R.J., Osmani S.A.,
RA DeSouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R.,
RA Yarden O., Plamann S., Seiler S., Dunlap J., Radford A., Aramayo R.,
RA Natvig D.O., Alex L.A., Mannhaupt G., Ebbole D.J., Freitag M.,
RA Paulsen I., Sachs M.S., Lander E.S., Nuebaum C., Birren B.;
RT "The Genome Sequence of the Filamentous Fungus Neurospora crassa."
RT Nature 0:0-0(2003).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.;
RL Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC German Neurospora genome project;
RA Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.

```

```

CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABX01000006; EAA36182.1; -; Genomic_DNA.
DR EMBL; BX844626; CAE76263.1; -; Genomic_DNA.
DR GO; GO:0008757; F:S-adenosylmethionine-dependent methyltransf. .; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000051; SAM bind.
KW Hypothetical protein; Methyltransferase; Transferase.
SQ SEQUENCE 908 AA; 101376 MW; C20AC06831AA4154 CRC64;

Alignment Scores:
Pred. No.: 9,76e-36 Length: 908
Score: 850.50 Matches: 222
Percent Similarity: 44.7% Conservative: 105
Best Local Similarity: 30.4% Mismatches: 235
Query Match: 23.5% Indels: 169
DB: 2 Gaps: 18

US-10-620-914-44 (1-1947) x Q7SGY5_NEUCR (1-908)
QY 205 CGCTCGAACCCTCATCTGGTGTGACCTGGTGGCACTGGGAGAAATGTCGATATGATG 264
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 178 ArgArgLysProIleTyrValAspValGlyGlyThrGlyTyrPheIleGluAlaMet 197
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 265 GCTGATTACATCGACCTGGCGAAG---TTCAAGTCCATCTACGTGTGCGACCTGTGCCAC 321
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 198 AlaLysPheValAsnValSerGluPheLysThrValTyrLeuValAspPheSerPro 217
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 322 TCCTGTGGAGTGGCCAAAGAAAGAGCGGAGCGGAGCGGAGAGAGATGTCAGGTC 381
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 218 SerLeuCysGluValAlaArgLysArgPheAlaArgLeuGlyTyrGluAsnValArgVal 237
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 382 GTGGAGCGGACGCTTGGCAATTTGCG-----CCCCCTGAGGGGACC 423
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 238 IleCysThrAspAlaArgLysPheArgLeuGluAspTyrGluAspValAspGluGlyGlu 257
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 423 -----423
Db 258 SerGlySerGlyAspSerSerProSerLeuSerGlyTyrTyrTyrGlyGluThrLysProGly 277
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 424 -----GGAGCGCTCATCACCCTTCTCTACTCGCTCAGCATGATCCACCTTC 471
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 278 ArgHisAlaGlyAlaGluLeuIleThrMetTyrSerLeuSerMetMetProAspTyr 297
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 472 CACAACTCATCGACCGAGGCTTGTCTACTCTGCCAAGCGGCTGTGGGGTGGTCC 531
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 298 PheSerIleLeuAspSerLeuGluSerLeuLeuAlaProHisGlyLeuIleAlaValVal 317
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 532 GACTTCTTACCTGAGCGGCAAGTACGAC-----CTGCCC 564
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 318 AspPheTyrAlaGlnSerLysValAspPheThrPheArgAsnTyrThrGlyGlyLeuMet 337
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 565 CTGGCGCAGATGCCCTGGTGGCGCGCTTCTCTGGCGATCGATCTTCGATCGACAC 624
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 338 AsnArgHisValGlyTyrPheAlaArgAsnPheTyrArgSerTyrPheAspAlaAspArg 357
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 625 ATTGACATCGGCGCGCGCTTCTACTCGAGCAGAGCTGGAGCGCGCTGG---681
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 358 ValSerLeuGluProAlaArgAspTyrLeuGluTyrArgPheGlyThrValLeuThr 377
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 682 -----GAGCAGAACACCCAGGTTTCGATCCCTACCTGCGCGTGG-----CTG 723
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 378 ValAsnAlaArgAsnThrLeuGlyAlaIleProTyrTyrIleTyrLeuGlyCysLeu 397
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 724 CGCGCCCCCTACTACGTG-----741
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 398 LysLysProPheSerThrSerSerLeuProHisGluIleValGluHisIleAspAlaIle 417
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 741 -----741
Db 418 AlaThrGluSerProArgSerSerProArgLeuValGlyLysHisSerSerSerAlaThr 437
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```



QY 742 -----TGGATTGGCCGC----- 753  
 Db 438 AsnAlaLeuAlaPheAlaValGlyArgThrAlaProGluMetArgSerIysAlaPheAsn 457  
 QY 754 -----CTGCCAGCGTTGGCCAGCGCCCTGCAC 780  
 Db 458 ThrAlaIleGluAsnIleSerAlaAsnLeuProLeuProSerPhePheTyrGlnAsnHis 477  
 QY 781 -----GAGGAGCGGTGGAGCGCGCCCATGTCTCCGCCACCTTC 822  
 Db 478 HisTyrArgIleTyrTyrAspAspGlnLeuProIysHisThrGlnPheAsnAspGluTyr 497  
 QY 823 CTGTACAGCGTGTGGAGAGACCCCGACCGGATATGAGGTGATGAGATCAACCC 882  
 Db 498 IleTyrAlaPheThrTyrGluAspSerArgValAspArgGluLeuLeuAsnLeuGlyPro 517  
 QY 883 AAGCACAGCTGTGACCTGTAGCGGGCTGTCAATCCCTGAACCTGTGGTGCAG 942  
 Db 518 AspAspValValLeuAlaIleThrSerAlaGlyAspAsnIleLeuSerTyrLeuMetGln 537  
 QY 943 GGGCGCGCGAGGTGTGTGGTGGACTGCAACCGCGCGAGTGGCGCTTCTGGAGCTG 1002  
 Db 538 SerProAlaArgValHisAlaIleAspLeuAsnProAlaGlnAsnHisLeuLeuGluLeu 557  
 QY 1003 AAGAAGTGGCCATTACAGAGCTGGAGTTTGAGAGCTGTGGCAGCTGTGGCGAGGC 1062  
 Db 558 LysValAlaSerPheThrThrLeuAspTyrProAspValTyrIysIlePheGlyGluGly 577  
 QY 1063 GTCCACCGCGCATTTGAGGAGCTGTACGAGAGAGCTGGCGCTTCTGTGCGCAACC 1122  
 Db 578 LysHisProAspPheArgSerLeuLeuIleSerLysLeuSerProHisLeuSerGlyArg 597  
 QY 1123 AGGCACAACTTCTGGTCCAGCGCTCTGTACTTCCAG-----CACGGCGCTG 1170  
 Db 598 AlaPheGlnTyrTrpLeuSerAsnAlaHisIlePheThrAspProAlaGlyArgGlyLeu 617  
 QY 1171 TACTACAGCGCGGATGGCAAGCTGTGTGGTGTGCAGTGCCTGGCGGTGGTGTG 1230  
 Db 618 TyrAspThrGlyGlySerArgTyrAlaIleArgPhePheArgTyrIleSerThrLeuPhe 637  
 QY 1231 GGACTGGCAAGCCTCAAGCGCTCGCCCAAGCGCCACATGAGGAGAGCGCGCT 1290  
 Db 638 PheCysArgSerAlaValArgArgLeuLeuSerThrProThrLeuGluGlyGlnArgSer 657  
 QY 1291 CTGTGGGACAGCAACTCTCATCTCTCTCGTGAAGAGCGCGCGCTGTGTGTG 1350  
 Db 658 IleTyrHisThrIysIle-----ArgProCys----- 666  
 QY 1351 CTGTTCTCAAGTCTGTGAGC---CTGTGCTCTTCAACAGCGCGCTGTGTGTGCTGCGC 1407  
 Db 667 LeuLeuAsnArgPheValAsnGlyLeuValLeuSerSerAspAlaPheLeuTyrSerAla 686  
 QY 1408 GGGCGGTGCGGGCAGCAGTACGCGCTGATCAAGCGGAC----- 1449  
 Db 687 LeuGlyValProIysAsnGlnValAlaMetIleGluAlaAspTyrHisArgArgSerIle 706  
 QY 1450 -----GGCATCCCATTTGAGAAC 1467  
 Db 707 SerSerSerThrThrProSerSerLysGluLysProSerArgAlaGluAlaLeuHis 726  
 QY 1468 TACATCGCGCGCACATGAGCGGTGGGAGAACTCGCACGTGGCGAGCAAGCAACTAC 1527  
 Db 727 TyrThrThrSerThrLeuAspProValLeuSerThrSerHisLeuAlaSerAspAsnPro 746  
 QY 1528 TTCTACTACAACTGCTCAGCGCAAGTCTCTGGCGCAACTGCGCCACCTACCTGCGC 1587  
 Db 747 TyrTyrLeuValCysValLeuGlyGlnTyrThrArgGlnCysHisProAspTyrLeuSer 766  
 QY 1588 GAGCGCGCTTCCGCCCTTCAAGAGT---GGCGTGTGGCAACCTACCGCTCTCCACC 1644  
 Db 767 ProAlaAlaHisSerIleLeuSerAlaProGlyAlaPheAspGlyLeuArgIleHisThr 786  
 QY 1645 AACTTCTTATGGAGGAGCTC-----AAAGCGCGCACCTTACACCAAGGTGATTCTG 1695

Db 787 AspGluIleGlnGluValLeuAlaArgPheGlnProGlyThrLeuThrValAlaVal 806  
 QY 1696 ATGGACACAGCTGAGCTGGCTGGATATGCC----- 1725  
 Db 807 MetAspSerMetAspTyrPheAspProProSerProGluGluLysGluGlyArgGly 826  
 QY 1726 GTGGCCACAGAGCTGGCGGAGTCCCTGGCCAGCAGTGTTCGCGCGCGCATCGTCATC 1785  
 Db 827 LysAlaArgGluGlnValArgArgLeuAsnArgAlaLeuLysValGlyGlyValLeu 846  
 QY 1786 TGGCGCTTCCGCTCCCTCAGCCCGCTACCGCGAGCTGATCCAGAGCGCGGCTTCGAC 1845  
 Db 847 LeuArgSerAlaGlyValGluProTyrTyrValArgValPheValGluGluGlyPheGly 866  
 QY 1846 GTGCGCTGTCATC-----CGCGCGCCACTCAGGCTACATGAGCCGCGTC 1890  
 Db 867 AlaArgArgValGlyCysArgGluSerGlyArgGlyAspGlnGluCysIleAspArgVal 886  
 QY 1891 AACATGTACAGCTCTTCTACATGCGCCCGCG 1923  
 Db 887 AsnMetTyrAlaSerCysTrpIleLeuGluLys 897  
 RESULT 11  
 Q5KX9\_CRYNE PRELIMINARY; PRT; 797 AA.  
 ID O5KX9\_CRYNE PRELIMINARY; PRT; 797 AA.  
 AC O5KX9;  
 DT 10-MAY-2005 (TrEMBLrel. 30, Created)  
 DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)  
 DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)  
 DE Hypothetical protein.  
 GN ORFNames=CNE01970;  
 OS Cryptococcus neoformans var. neoformans JEC21.  
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Heterobasidiomycetes;  
 OC Tremellomycetidae; Tremellales; Tremellaceae; Filobasidiella.  
 OX NCBI\_TaxID=214684;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=JEC21;  
 RA Loftus B., Amedeo P., Roncaglia P., Vamathevan J., Utterback T.,  
 RA Van Aken S., Fraser C.; to the EMBL/GenBank/DBJ databases.  
 RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RC STRAIN=JEC21;  
 RX PubMed=15653466; DOI=10.1126/science.1103773;  
 RA Loftus B.J., Fung E., Roncaglia P., Rowley D., Amedeo P., Bruno D.,  
 RA Vamathevan J., Miranda M., Anderson I.J., Fraser J.A., Allen J.E.,  
 RA Bosdet I.E., Brent M.R., Chiu R., Doering T.L., Donlin M.J.,  
 RA D'Souza C.A., Fox D.S., Grinberg V., Fu J., Fukushima M., Haas B.J.,  
 RA Huang J.C., Janbon G., Jones S.J.M., Koo H.L., Krzywinski M.I.,  
 RA Kwon-Chung K.J., Lengeler K.B., Maiti R., Marra M.A., Marra R.E.,  
 RA Mathewson C.A., Mitchell T.G., Pertea M., Riggs F.R., Salzberg S.L.,  
 RA Schein J.E., Shvartsbeyn A., Shin H., Shumway M., Specht C.A.,  
 RA Suh B.H., Tenney A., Utterback T.R., Wickes B.L., Wortman J.R.,  
 RA Wye N.H., Kronstad J.W., Lodge J.K., Heitman J., Davis R.W.,  
 RA Fraser C.M., Hyman R.W.;  
 RT "The genome of the basidiomycetous yeast and human pathogen  
 RT Cryptococcus neoformans.";  
 RL Science 307:1321-1324 (2005).  
 DR EMBL; AB017345; AAW43531.1; -; Genomic DNA.  
 KW Complete proteome; Hypothetical protein.  
 SQ SEQUENCE 797 AA; 90388 MW; FD1082B7EB2B240 CRC64;

Alignment Scores:  
 Pred. No.: 1,73e-35 Length: 797  
 Score: 845.50 Matches: 221  
 Percent Similarity: 42.8% Conservative: 104  
 Best Local Similarity: 29.1% Mismatches: 220  
 Query Match: 23.4% Indels: 215  
 DB: 2 Gaps: 21

US-10-620-914-44 (1-1947) x Q5KX9\_CRYNE (1-797)



```

Db 753 LysArgProTyrArgGlnArgPheGluAlaGlyLeuLysValGlnProIleAsp 772
QY 1861 ---CGGCCCACTCAGGCTACATGCGCGGTCAACATGTACAGCTCTCTTCTACATGCGC 1917
Db 773 IleArgGluAnGlnGluAlaIleAspArgValAsnMetTyrAlaSerPheTrpLysAla 792

RESULT 12
ID Q4X175 ASPFU PRELIMINARY; PRT; 831 AA.
AC Q4X175;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Hypothetical protein.
DE
GN ORFNames=Afu2g10950;
OS Aspergillus fumigatus Af293.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=330879;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Af293;
RA Nierman W., Pain A., Anderson M.J., Wortman J., Kim H.Stanley.,
RA Arroya J., Berriman M., Abe K., Archer D.B., Bermejo C., Bennett J.,
RA Bowyer P., Chen D., Collins M., Coulson R., Davies R., Dyer P.S.,
RA Farman M., Fedorova N., Fedorova N., Feldblyum T.V., Fischer R.,
RA Posner N., Fraser A., Garcia J.L., Garcia M.J., Goble A.,
RA Goldman G.H., Gomi K., Griffith-Jones S., Gwilliam R., Haas B.,
RA Haas H., Harris D., Horiuchi H., Huang J., Humphrey S., Jimenez J.,
RA Keller N., Khouri H., Kitamoto K., Kobayashi T., Kulkarni R.,
RA Kumaai T., Latton A., Latton J.-P., Li W., Lord A., Lu C.,
RA Majoros W.H., May G.S., Miller B.L., Mohamoud Y., Molina M., Monod M.,
RA Mouyna I., Mulligan S., Murphy L., O'Neil S., Paulsen I.,
RA Penava M.A., Pertea M., Price C., Pritchard B.L., Quail M.A.,
RA Rabinowitz E., Rawlins N., Rajandream M.-A., Reichard U.,
RA Renauld H., Robson G.D., Rodriguez de Cordoba S., Rodriguez-Pena J.M.,
RA Ronning C.M., Rutter S., Salzberg S.L., Sanchez M.,
RA Sanchez-Perrero J.C., Saunders D., Seeger K., Squares R., Squares S.,
RA Takeuchi M., Tekai F., Turner G., Vazquez de Aldana C.R., Weidman J.,
RA White O., Woodward J., Yu J.-H., Fraser C., Galagan J.E., Asai K.,
RA Machida M., Hall N., Barrell B., Denning D.W.,
RT "Genomic sequence of the pathogenic and allergenic filamentous fungus
RT Aspergillus fumigatus."
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC DR EMBL; AAHF01000001; EAL93390.1; -; Genomic_DNA.
CC DR InterPro; IPR000051; SAM_bd.
CC KW Hypothetical protein; Methyltransferase; Transferase.
CC SQ SEQUENCE 831 AA; 94737 MW; D3601019DB480CB4 CRC64;

Alignment Scores:
Pred. No.: 1.81e-33 Length: 831
Score: 806.50 Matches: 224
Percent Similarity: 42.4% Conservative: 95
Best Local Similarity: 29.7% Mismatches: 257
Query Match: 22.3% Indels: 177
DB: 2 Gaps: 20

US-10-620-914-44 (1-1947) x Q4X175_ASPFU (1-831)
QY 124 AAGNAGGC---GATGATCAGCTGCTCGCTGAGAGCTTCTACGGGCCCGGCGCT 180
Db 69 LysLeuGlyLeuAsnGlyGlnGlnAspAlaLeuGluSerPheTyrLysThrGlnAlaGly 88
QY 181 GCCTTT----- 186
Db 89 ValTyrAspAlaThrArgLysArgLeuLeuCyseGlyArgGluAspMetLeuGlyLeuVal 108
QY 187 GTCGCGCGCTG-----GCCGAGCGCTCGAACCTCATCTGG 222
|||||:|||||

```

```

Db 109 AlaAlaGlnLeuLysTyrLysValGluAsnLysGluLeuGlnAlaGlyLysAlaIleTrp 128
QY 223 GTTGACCTGGGTGCTGGCACTGGGAGAATGTGATATGATGCTGATGATCATCATCGACTG 282
Db 129 ValAsp-----AspSerTyrAsnIleGluAlaMetAlaSerPheLeuProVal 144
QY 283 GCAGAGTTCAAGTCC---ATCTACGTGGTGCACCTGTGCACCTGCTGTCGAGGTGGCC 339
Db 145 ThrGlnPhePheSerHisValTyrLeuValAspLeuSerProSerLeuCyseGluValAla 164
QY 340 AAGAAGAGCGGGAAGCGCTCGAAGAGTCCAGGTGCTGAGGCGCGGCTTGC 399
Db 165 ArgGlnArgPheGluArgLeuGlyTrpLysAsnValThrValValCyseGlnAspAlaArg 184
QY 400 CAATT-----GCGGCCCTCGAGGGCACC 423
Db 185 SerPheArgLeuProHisGluLysValAspProArgAlaProAlaProSerThrAlaGly 204
QY 424 GCAGCGCTCATCACCCTTCTCTACTCGTCACGATGATTCACCGTTCCACACGTCATC 483
Db 205 AlaAspLeuIleThrMetSerTyrSerLeuSerMetIleProAspTyrTyrSerValVal 224
QY 484 GACCAGCTTGTCTGCTACCTGTCCTCCAGACGCGCTGGTGGCGCTTGGCGACTTCTAGTG 543
Db 225 AspSerLeuThrAspLeuLeuArgAlaSerGlyIleLeuGlyValCyseAspPheTyrVal 244
QY 544 AGCGGCAAGTACGACTGCTGCGCTG-----CGCCAGATG 576
Db 245 GlnSerIleValAspValSerAlaArgAsnTyrIleGlyAlaPheAsnArgHisVal 264
QY 577 CCTGTGTCGCGCGCTTCTTCTGGCGATCGATCTTCGACATCGACACATCATCATCGCG 636
Db 265 AsnTrpLeuGlyArgAlaPheTrpArgAlaTrpPheAspAlaAspArgValSerLeuGlu 284
QY 637 CCGAGCGCGCGCTACCTGAGAGAGAGTGGAGCGGTGGGAGGAGACACCCAG 696
Db 285 AlaAlaArgArgAspTyrLeuGluTyrArgPheGlyThrValIleSerAlaSerGluArg 304
QY 697 -----GGTTCGATCCCTACGTGCGGTGCGCGCTGCTGCTGCTGCTGCTGCTGCTG 744
Db 305 AsnTyrLeuLeuGlyGlyle-----ProTyrTyrIlePhe 316
QY 745 ATTGCGCGC----- 753
Db 317 ValGlyArgHisLysAsnLeuAlaSerAsnLeuSerGlyGlnGlnThrIleGluMetLeu 336
QY 753 ----- 753
Db 337 AspAlaSerPheThrGluSerProTyrLeuSerProAlaAsnHisThrLysGluMetGlu 356
QY 753 ----- 753
Db 357 LysAlaMetGluArgAsnAlaGlnGluIleArgSerLysAlaTyrGluSerAlaValIle 376
QY 754 -----CTGCCAGCGTTGGCCACGCCCTGCACGAGGAGCGCGTG 792
Db 377 AsnLeuSerAlaAsnLeuProLeuProSerPheTyrGlnAsnHisHisCysArgIle 396
QY 793 -----GAGCGCGCGCCCATGTTCCCGCCCGCCCTTCTGTCACGCGAG 834
Db 397 PheTyrAsnAspLeuLeuProLysHisThrGlnPheLysAsnGluTyrIleTyrAlaPhe 416
QY 835 TCGTGGGAGGAGCCCGGAGCGGATATGAGGTGATGAGATCAACCCCAAGGACACGGTG 894
Db 417 AsnTrpGluAspProArgValAspHisArgLeuLeuAspIleLysArgAspAspValIle 436
QY 895 CTGACCTGACTAGCGCGCTGCAATGCCCTGGAACCTGCTGTGTGAGCGGGCGCGCGCAG 954
Db 437 LeuAlaIleThrSerAlaGlyAspAsnIleLeuAspTyrLeuGlnLysSerProArgArg 456
QY 955 GTGCTGTCGTGAGCTGCAACCCCGCGAGTGGCGCGCTTCTGAGCTGAAGAGGTGGCC 1014
Db 457 ValHisAlaValAspLeuAsnProAsnGlnAsnHisLeuLeuLeuLysValAlaSer 476
|||||:|||||

```





OX NCBI\_TaxID=117;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC  
 RA MEDLINE=22735913; PubMed=12835416; DOI=10.1073/pnas.143143100;  
 RX Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,  
 RA Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,  
 RA Schleener H., Amann R., Reinhardt R.;  
 RT "Complete genome sequence of the marine planctomycete *Pirellula* sp.  
 RT strain 1.1";  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303 (2003).  
 DR EMBL; BX294133; CAD71533.1; -; Genomic DNA.  
 KW Complete proteome; Hypothetical protein.  
 SQ SEQUENCE 415 AA; 47529 MW; 88D3194F7446D252 CRC64;

Alignment Scores:  
 Pred. No.: 9,09e-16 Length: 415  
 Score: 463.00 Matches: 124  
 Percent Similarity: 48.1% Conservative: 79  
 Best Local Similarity: 29.4% Mismatches: 158  
 Query Match: 12.8% Indels: 62  
 DB: 2 Gaps: 13

US-10-620-914-44 (1-1947) x Q7UYV8\_RHOBA (1-415)  
 QY 742 TGGATTGGCGCGCTCCAGCGGTTGGCCACGCTCGACGAGGAGCGCGTGGAGCGCGC 801  
 DB |||||  
 5 TrpLeuGlyAsn-----LysCysPheLysValValHisGlnLysAsnLeu----- 19  
 QY 802 CCCATGTTCCCGGCCACTTCTGTACACGAGTCGTGGGAGGACCCCGAGCGGATATG 861  
 DB |||||  
 20 -----ValTyrAsnThrCysTrpGluAspProArgLysAspArg 32  
 QY 862 GAGTGTAGATCAACCCAGGACACGGTGTGACCTGACTAGCGCGGCTGCAAT 921  
 DB ::::|  
 33 GlnAlaLeuSerLeuGlyProAspAspSerValLeuValIleThrSerAlaGlyCysAsn 52  
 QY 922 GCCTGAACTGCTGGTGGAGGGCGCGCGAGTGTGCGTGGAGTCGCAACCCCGCG 981  
 DB |||||  
 53 AlaLeuAspTyrAlaLeuGlnAlaProLysSerValHisAlaValAspMetAsnProLeu 72  
 QY 982 CAGTGGCGCTTCTGGAGCTGAAGAAGGTGGCCATTCAGCAGCTGGAGTTTGAGGACGTG 1041  
 DB |||||  
 73 GlnAsnAlaLeuLeuGluLeuLysCysAlaSerIleArgCysLeuThrTyrAspAspPhe 92  
 QY 1042 TGGCAGCTGTTCCGCGAGGGGTCCACCGCGCATTCAGGAGCTGTACGAGAGCAAGACTG 1101  
 DB ::::|  
 93 PheAsnValPheGlyArgGlyTyrHisProAlaTrpGlnSerLeuTyrHisLysHisVal 112  
 QY 1102 GCGCCCTTCTGTCGCAACACGACGACCACTTCTGGTCCAGCGCTCTGTACTTTC--- 1158  
 DB |||||  
 113 ArgGlyAlaLeuThrProAspThrArgAlaIleTrpAspArgLysLeuAspPheAsp 132  
 QY 1159 -----CAGCAGCGCTCTACTACAGCGCGCATGGCAAGCTGTGCTGGGTG 1206  
 DB ::::|  
 133 GlyThrSerArgArgLysSerPheTyrPheArgGlyThrSerGlyLeuPheAlaTrpLeu 152  
 QY 1207 CTGAG---TGCTGGCGGTGCTGCTGGAGTGGCGAAGACCGTCAAGCGCTCGCCAC 1263  
 DB ::::|  
 153 ValAsnGlyTyrLeuLysArgProAlaGlyLeuArgGluAlaIleAlaGluLeuGln 172  
 QY 1264 GCGCCCAACATGGAGGAGCGCGCTGCTGGGACAGCAACATGCTCCACTTCCTGCT 1323  
 DB |||||  
 173 AlaGluSerValAspGlnArgGluIleTyrGlnLysArgGluIleAsnArgLeuLeu 192  
 QY 1324 AAGAACGGGCGCAAGCGCTGGTGTGCTGTCGTCGAGCTGAGCGCTGGCTCTTC 1383  
 DB ::::|  
 193 TrpSer-----LysProLeuArgTrpAlaLeuArgArgAspThrThrLeuAlaMetLeu 210  
 QY 1384 -----AACAGCGCTGCTGTTGCTGGCGGGCGGCTGCCGGC 1422  
 DB |||||  
 211 GlyValProArgSerGlnArgAsnGlnLeuAspGlnCysTyrProGlyGlyGlyGly 230

QY 1423 AAGCAGTACGCGCTGATGATCAAGCGGACGCGCATCCCATTTGAGAACTACATCGCGCGCAC 1482  
 DB ::|||  
 231 -----PheIleGlnAspArg 235  
 QY 1483 ATGGACGGCGTGGCGGAGAACTCGCAGTGGCGAAGCAGAACTACTTCTACTACAACCTGC 1542  
 DB ::|||  
 236 IleGluAlaValPheLysThrLeuProLeuArg---AspAsnTyrPheTrpArgValTyr 254  
 QY 1543 CTCACCGGCAAGTTCCTGCGGACAACTGCCCCACCTACCTCGCGGAGGGCGCTTCGCGC 1602  
 DB |||||  
 255 LeuThrGlySerTyrThrProAspCysCysProGluTyrLeuLysGluGluAsnPheThr 274  
 QY 1603 ACCCTCAAGAGTGGCGTGGGACCACTGACCGTCTCCACCACTTC-----TTC 1653  
 DB |||||  
 275 GluLeuLysAsnGlyLeuValAspCysValGlnThrHisThrAsnThrValGluGlyPhe 294  
 QY 1654 ATGGAGAGCTCAAGCGGCGCCTACACCAAGTGTCTGTGACCACTGAGCACTGGACTGG 1713  
 DB ::|||  
 295 LeuThrGlu---HisAspGlnProIleSerArgPheValLeuLeuAspHisMetAspTrp 313  
 QY 1714 CTGGATATGCCCGTGGCGGACGAGTGGCGGAGTGTGCGCCCAAGCAGGTTCGCGCGGCG 1773  
 DB |||||  
 314 Leu-TyrAspArgTyrProGluLeuLeuAlaSerGluTrpGlnSerIleLeuAsnArgAl 333  
 QY 1774 GGCATC-----GTCATCTGGCGCTCGCGCTCCCTCCAGCGCGCTACGCGGAG-- 1821  
 DB |||||  
 333 aThrSerAspAlaArgValLeuTrpArgSerAlaAlaLeuAlaValAspPheValAsnPr 353  
 QY 1822 ----CTGATCCAGAGCGGCGCTTCGACGTGCGCTG----- 1854  
 DB ::|||  
 353 oLeuMetLeuGlnHisAsnGlySerGlnValGlnLeuGlyAspLeuLeuHisTyrHisAs 373  
 QY 1855 -ATCCGCGCGGCGCACTCAGGCTGACATGACCGCGCTCAACATGTACAGCTCTCTCTACAT 1913  
 DB ::|||  
 373 pGluLeuAlaSerSerLeuHisAlaArgAspArgValHisThrTyrGlySerPheTyrIle 393  
 QY 1914 GGCC 1917  
 DB ::|||  
 393 eAla 394

RESULT 15  
 Q4NR00\_9DELTA PRELIMINARY; PRT; 767 AA.  
 AC Q4NR00;  
 DT 13-SEP-2005 (T-EMBLrel. 31, Created)  
 DT 13-SEP-2005 (T-EMBLrel. 31, Last sequence update)  
 DT 13-SEP-2005 (T-EMBLrel. 31, Last annotation update)  
 DE Liga.  
 GN ORFNames=AdehdRAFT\_0943;  
 OS Anaeromyxobacter dehalogenans 2CP-C.  
 OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;  
 OC Cytophactereae; Myxococcaceae; Anaeromyxobacter.  
 OX NCBI\_TaxID=290397;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=2CP-C;  
 RG US DOE Joint Genome Institute (JGI-PGF);  
 RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,  
 RA Hammon N., Israni S., Pittluck S., Richardson P.;  
 RT "Sequencing of the draft genome assembly of Anaeromyxobacter  
 RT dehalogenans 2CP-C.";  
 RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=2CP-C;  
 RG US DOE Joint Genome Institute (JGI-ORNL);  
 RA Larimer F., Land M.;  
 RT "Annotation of the draft genome assembly of Anaeromyxobacter  
 RT dehalogenans 2CP-C.";  
 RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.  
 CC -!- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.



QY	1138	ACCAGAAGTTGTGGC-----TGGTTTTGGACAGGAAGGGCGCCAGCTTCTTCTCTGTA	1081
Db	303	-----AlaCyseGlyArgArgTrpGlyAlaAaATrp-----ProAlatPrThrAlaAarg	318
QY	1084	GCTCTCAATGCGCGGTGCAGCCCTCGCCGAACAGCTGCCACACGTCCTCAAACCTCCA	1025
Db	319	AlaProThrArgCyseGlyAlaAargSerProAargThrIleAlaSerArgArCyseAlaGly	338
QY	1024	GCTGCTGAATGGCCACTTCTTCAGCTCCAGAAAGCCGACCTCGCGGGGTTCAGT---	968
Db	339	ArgProGlyTrp-----ArgAlaAlaAargAgAlaAargCyseAlaSerGlyCyserArg	356
QY	967	-----CCACCGCACACACTGGC-----	950
Db	357	SerArgArgSerAlaCyseSerAlaAargTrpArgAlaThrProArgArgGlyProGlyPro	376
QY	949	-----CGGCCCTCGACCA-	935
Db	377	CysAlaSerAlaAargSerProSerProAargArgArgSerThrArgTrpThrGlyTrpThr	396
QY	934	--GCAGGTTCAGGCATTGCAGCCCGCTAGCTCAGGGTCAGCACCCGCTGCTCT--	884
Db	397	ThrArgGlyProGlyArgCyseGlyProAargAlaSerProAlaGlyProProProSer	416
QY	883	-----TGGGTGTGATCTCATCACT--	848
Db	417	ArgArgSerGlyThrSerProAargProAargProAlaGlyArgSerSerArgProAargSer	436
QY	847	GGTCTCCCCAGACTCGGTG--	812
Db	437	GlyProProAargAlaCyseProCyseGlyThrArgThrArgCyseSerArgTrpPro	456
QY	811	GGAACTAGCGCGCCCTCCA-----CGCGCTCTCTGCT--	779
Db	457	ArgProAargAlaCyseProProAargTrpAargSerHisAlaProAargaspValProPro	476
QY	779	-----	779
Db	477	ArgAspArgProAlaHisAlaAlaAalaAargHisGlyAlaProAlaHisArgGlyGlyAla	496
QY	778	---GCAGGGCGTGGCCAAAGC-----TGGCGAGCGGCCCAATCCACACGT	737
Db	497	GlyGlyGlyArgGlyAlaAargValProProLeuAlaAlaHisGlyProAlaAlaAala	516
QY	736	AGTAGGGGCGCGCACCCACCGCA-----CGTAGGGGATCGAAC-----	698
Db	517	GluaAlaGlyAspGlyAlaGlyAlaAappProAlaValAargArgGlyAspHisHisHis	536
QY	697	-----CCTGGGTGTTCTGCTCCC	680
Db	537	GlyAspArgArgGlnAargGlyGluGlyValAargProGlyGlyProGlyGlnLeuAargPro	556
QY	679	ACAGCGCTCCAGCTTCTGCTCCAGTAGCGC-----	647
Db	557	ValProLeuAargAlaGluGlyProAlaAargAargGlyGlyAlaValProHisHisAargPro	576
QY	646	-----GGCGCTCGGGCGCATGTCATGTTGTCGATGTCGAAGATCGATCGCCAGAAGA	593
Db	577	GlyHisGlyValAargSerAargaspAlaGlyAlaAargAlaHisAargAargAlaAargGlyVal	596
QY	592	AACGGCGCACCAAGGCA-----TCTGGCGAGGGCGAGTCTGCTACT	551
Db	597	GlyGlyAlaGlyAargAlaAargProAlaProGlyAargArgGlyAlaGlyAargGlyAargAla	616
QY	550	TGCGCGTCAAGTAGAGTGGCAACGCCCAACGAGCGGCTCTGGGACAGGTACGACGACAG	491
Db	617	ArgGlyAargAargAlaAargAargAargAlaAargGlyAargGlyGlyAlaAlaAala	635
QY	490	CCTGTTGATGACGTTGTGGAACGGTGGAAATCATCTGTAGCGAGTAGGACAAGCTGATGA	431
Db	636	AlaAargArg-----ArgValAlaGluValAlaValAargAargSerArgAlaAla	651
QY	430	GCCTCGCGGTGCCCTCAGGGGGCGCAAAATGGCAAGCGTCGGCTCCA-----	383

```

Db      652 IleGlyArgTyrProArgArgAlaAlaProArgArgAlaProAlaAlaArgAsp 671
Qy      382 CGACCTGGACATTCTTCCAGCCCT-----TGGCCTTCGCCTTCT 344
Db      672 ArgArgSerCysAlaValAlaProAlaThrAlaGluArgAlaIleTyrProGlyAsnAla 691
Qy      343 TCTTGGCCACTCGCACAGCGAGTGGCACAGGTGCA---CCACGTAGATGGACTTGA 287
Db      692 ProArgArgProGluSerArgProAlaGluGlyArgSerProArgProLeuAlaLeuGly 711
Qy      286 TCGCCAGGTGATGATTAATCAGCCATCATATCGACATTCTCCCGAGTCCACCCAGGT 227
Db      712 HisProGlySerCys-----HisPro--- 718
Qy      226 CAACCCAGATGAGGTTTCGAGCGTTCGGCCAGCGGCGAGCAAGGCGCGCTTGGGGCC 167
Db      719 -----AlaArgAspArgGlyGlnCysSerGlnMetProValArg 731
Qy      166 CGTAGAGCTTCCAGCGGAGCAGCGTGAT 137
Db      732 AlaArgGluSerGlyArgAspAsnValAsp 741

```

Search completed: March 14, 2006, 01:44:31  
 Job time : 291 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 13, 2006, 21:32:22 ; Search time 7964 Seconds  
(without alignments)  
11438.278 Million cell updates/sec

Title: US-10-620-914-44

Perfect score: 1947  
Sequence: 1 atggggtcgggtcgtgacgg.....gogccaaagagacaactaa 1947

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_hic:\*  
5: gb\_est4:\*  
6: gb\_est5:\*  
7: gb\_est6:\*  
8: gb\_est7:\*  
9: gb\_gss1:\*  
10: gb\_gss2:\*  
11: gb\_gss3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	725	37.2	743	5	BU651274	BU651274 1112092D1
2	556	28.6	556	1	AV627635	AV627635 AV627635
3	470.4	24.2	475	1	AV644266	AV644266 AV644266
4	463.8	23.8	467	1	AV629528	AV629528 AV629528
5	414	21.3	532	1	AV628087	AV628087 AV628087
6	391	20.1	391	1	AV626756	AV626756 AV626756
7	379	19.5	379	3	BP086130	BP086130 BP086130
8	354	18.2	354	1	AV631627	AV631627 AV631627
9	352	18.1	352	3	BP093875	BP093875 BP093875
10	336	17.3	454	3	BP093208	BP093208 BP093208
11	319	16.4	508	1	AV641303	AV641303 AV641303
12	310	15.9	511	1	AV626139	AV626139 AV626139
13	287	14.7	478	1	AV624589	AV624589 AV624589
14	286	14.7	480	1	AV643824	AV643824 AV643824
15	267	13.7	466	1	AV628989	AV628989 AV628989
16	262	13.5	262	1	AV630897	AV630897 AV630897
17	255	13.1	461	5	BQ824076	BQ824076 1030115FO
18	250	12.8	447	1	AV642307	AV642307 AV642307
19	238	12.2	250	3	BP089550	BP089550 BP089550
20	189	9.7	398	3	BP093467	BP093467 BP093467
21	188.8	9.7	287	1	AV640882	AV640882 AV640882
22	188.8	9.7	362	3	BP096956	BP096956 BP096956

23	188.8	9.7	366	3	BP098395	BP098395 BP098395
24	188.8	9.7	377	3	BP097104	BP097104 BP097104
25	188.8	9.7	388	3	BP095133	BP095133 BP095133
26	174.4	9.0	416	6	CF556716	CF556716 1115020E0
27	172.8	8.9	176	1	AV628222	AV628222 AV628222
28	143	7.3	585	3	BI721528	BI721528 1031056E0
29	107.6	5.5	973	11	CNS07804	AL433370 T7 end of
30	99.2	5.1	433	1	AW561229	AW561229 ga77b12.y
C 31	88	4.5	838	11	CNS076PM	AL431696 T3 end of
C 32	85.6	4.4	1060	10	CW929808	CW929808 EDCBY94TF
33	82	4.2	850	11	CNS06VGM	AL417136 T3 end of
34	73.6	3.8	829	8	CV772865	CV772865 FGNS06726
35	71	3.6	488	1	AV624788	AV624788 AV624788
36	69.8	3.6	1895	11	DQ035307	DQ035307 Pan trogl
37	69.2	3.6	841	3	BI949299	BI949299 HVSME1001
C 38	67.4	3.5	963	11	CNS071FB	AL424845 T7 end of
C 39	67	3.4	301	2	BF072578	BF072578 NCSM3D4T3
C 40	67	3.4	417	10	CZ775539	CZ775539 OC_Ba013
C 41	66	3.4	688	10	CZ735917	CZ735917 OC_Ba007
42	65.6	3.4	792	7	CO894488	CO894488 Bovgen 22
43	64.8	3.3	528	2	BG464204	BG464204 EMI_70_H0
44	64.8	3.3	548	1	AW678218	AW678218 WSI_14_E0
45	64.8	3.3	553	2	BG240360	BG240360 OVI_29_A1

#### ALIGNMENTS

RESULT 1  
BU651274  
LOCUS  
DEFINITION  
1112092D1.y1 C. reinhardtii CC-1690 (mt+), CC-1691 (mt-), Gamete (normalized), Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.

ACCESSION  
BU651274  
VERSION  
BU651274.1 GI:23363454

KEYWORDS  
Chlamydomonas reinhardtii

SOURCE  
Chlamydomonas reinhardtii

ORGANISM  
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.

REFERENCE  
1 (bases 1 to 743)  
Grossman, A., Chang, C.-W., Davies, J., Harris, E., Hauser, C., Lefebvre, P., McDermott, J.P., Shrager, J., Silflow, C. and Stern, D.

AUTHORS  
Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants. Project: 1112

TITLE  
Unpublished (2002)

JOURNAL  
Contact: Charles Hauser

COMMENT  
DCMB Box 91000  
Duke University  
Durham, NC 27708-1000

Tel: 919 613 8159  
Fax: 919 613 8177  
Email: chauser@duke.edu.

Location/Qualifiers  
1. .743  
/organism="Chlamydomonas reinhardtii"

/mol\_type="mRNA"

/strain="21gr (CC-1690 wild type mt+)" & 6145c (CC-1691 wild type mt-)"

/db\_xref="taxon:3055"

/clone\_lib="C. reinhardtii CC-1690 (mt+), CC-1691 (mt-), Gamete (normalized), Lambda Zap II"

/note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2: XhoI; Gamete library was constructed by Hui Zhao, Min Lu, Jeffrey McDermott, William J. Shell and John Davies.

Strain 21gr cells (CC-1690; mating type plus) and strain 6145c cells (CC-1691; mating type minus) that had been growing on a light-dark cycle (13:11 L/D) in R-medium (Sager and Granick) were separately transferred into nitrogen-free medium at 8 hours into the light period.

PolyA mRNA was purified from each sample every 2 hours for

the next 18 hours. The mRNA was pooled and used for cDNA synthesis. The cDNA was directionally cloned into lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the lambda ZAP clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al., (1996) Genome Research 6: 791-806."

## ORIGIN

```

Query Match      37.2%; Score 725; DB 5; Length 743;
Best Local Similarity 99.7%; Pred. No. 6.5e-139;
Matches 736; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 681 GGAGCAGAACACCCAGGCTTCGATCCCTTAGTGCCTGCTGCGCGCCCTTACTACGT 740
    |||||
Db 7 GGAGCAGAACACCCAGGCTTCGATCCCTTAGTGCCTGCTGCGCGCCCTTACTACGT 66
    |||||
Qy 741 GTGATTTGGCGCGCTGCCAGCGTTGGCCACGCGCTTCACAGGAGGCGCTGGAGCGGCC 800
    |||||
Db 67 GTGATTTGGCGCGCTGCCAGCGTTGGCCACGCGCTTCACAGGAGGCGCTGGAGCGGCC 126
    |||||
Qy 801 GCCCATGTTCCCGCCCACTTCTGTACACGAGTGTGGAGGACCCCGAGCGGATAT 860
    |||||
Db 127 GCCCATGTTCCCGCCCACTTCTGTACACGAGTGTGGAGGACCCCGAGCGGATAT 186
    |||||
Qy 861 GGAGGTGATGAGATCAACCCCAAGGACAGGTGCTGACCTGACTAGCGCGCTGCCAA 920
    |||||
Db 187 GGAGGTGATGAGATCAACCCCAAGGACAGGTGCTGACCTGACTAGCGCGCTGCCAA 246
    |||||
Qy 921 TGCCCTCAACTGCTGTGAGGGGCGCGCGAGTGTGCTGCTGAGTGCACACCCCGC 980
    |||||
Db 247 TGCCCTCAACTGCTGTGAGGGGCGCGCGAGTGTGCTGCTGAGTGCACACCCCGC 306
    |||||
Qy 981 GCAGTCGCGCTTCTGAGCTGAAGAAGGTGGCCATTACAGCAGTGGAGTTGAGGACGT 1040
    |||||
Db 307 GCAGTCGCGCTTCTGAGCTGAAGAAGGTGGCCATTACAGCAGTGGAGTTGAGGACGT 366
    |||||
Qy 1041 GTGCGACCTGTTCCGCGAGGGCGTGACCCCGGATTTGAGGAGCTGTACGAGAAGACT 1100
    |||||
Db 367 GTGCGACCTGTTCCGCGAGGGCGTGACCCCGGATTTGAGGAGCTGTACGAGAAGACT 426
    |||||
Qy 1101 GGGCGCTTCTGTCGCAAAACAGCCCAACTTCTGCTCAAGCGCTCTGCTGTTACTTCCA 1160
    |||||
Db 427 GGGCGCTTCTGTCGCAAAACAGCCCAACTTCTGCTCAAGCGCTCTGCTGTTACTTCCA 486
    |||||
Qy 1161 GCAGCGCTGTACTACAGGCGCGCATGGGCAAGCTGTGCTGCTGCTGCTGCTGCTGCTG 1220
    |||||
Db 487 GCAGCGCTGTACTACAGGCGCGCATGGGCAAGCTGTGCTGCTGCTGCTGCTGCTGCTG 546
    |||||
Qy 1221 GGTGCTGCTGGGACTGGGCAAGCCGTCAAGCGCTTCGCAAGCGCGCCCAATGGAGGA 1280
    |||||
Db 547 GGTGCTGCTGGGACTGGGCAAGCCGTCAAGCGCTTCGCAAGCGCGCCCAATGGAGGA 606
    |||||
Qy 1281 GCAGCGCTGTGTCGAGCAGCAACATGCTCATCTTCTGTAAGAACGGGCGCCCAAGCC 1340
    |||||
Db 607 GCAGCGCTGTGTCGAGCAGCAACATGCTCATCTTCTGTAAGAACGGGCGCCCAAGCC 666
    |||||
Qy 1341 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1400
    |||||
Db 667 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 725
    |||||
Qy 1401 GTTCGGCGGGCGGCTGCC 1418
    |||||
Db 726 GTTCGGCGGGCGGCTGCC 743
    |||||

```

## RESULT 2

```

AV627635 AV627635 556 bp mRNA linear EST 15-DEC-2000
LOCUS AV627635 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas
DEFINITION reinhardtii cDNA clone LCL029g11_r 5', mRNA sequence.
ACCESSION AV627635

```

VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AV627635.1 GI:10790269

## EST.

Chlamydomonas reinhardtii

Chlamydomonas reinhardtii  
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;  
Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.

## REFERENCE

1 (bases 1 to 556)

## AUTHORS

Aamizu, E., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohyama, K.,

Nakamura, Y. and Tabata, S.

## TITLE

Adaptation of expressed sequence tags from low-CO2 and high-CO2

## JOURNAL

generated cells of Chlamydomonas reinhardtii

## PUBMED

11089912

## COMMENT

Contact: Erika Asamizu

The First Laboratory for Plant Gene Research

Kazusa DNA Research Institute

Yana 1532-3, Kisarazu, Chiba 292-0812, Japan

Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

## FEATURES

Location/Qualifiers

1..556

/organism="Chlamydomonas reinhardtii"

/mol\_type="mRNA"

/strain="C9"

/db\_xref="taxon:3055"

/clone\_lib="LCL029g11\_r"

/note="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:

XhoI; The cDNA library was constructed from cells cultured

in a carbon stress acclimated condition in which carbon

dioxide concentration in the bubbling gas was changed from

5% to 0.04%"

## ORIGIN

```

Query Match      28.6%; Score 556; DB 1; Length 556;
Best Local Similarity 100.0%; Pred. No. 4.5e-104;
Matches 556; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 284 CGAAGTTTCAAGTCCATCTAGCTGTCGACCTGTCGCACTCGCTGTGGAGGTGGCCAAAGA 343
    |||||
Db 1 CGAAGTTTCAAGTCCATCTAGCTGTCGACCTGTCGCACTCGCTGTGGAGGTGGCCAAAGA 60
    |||||
Qy 344 AGAAGCGAGGCGCAAGGGCTGGAAGAAATGCCAGGTCTGAGAGCGCAGCTTGGCAAT 403
    |||||
Db 61 AGAAGCGAGGCGCAAGGGCTGGAAGAAATGCCAGGTCTGAGAGCGCAGCTTGGCAAT 120
    |||||
Qy 404 TTGCGGCCCTGAGGGCACCAGCGCTCATCACTTCTCTACTCTGCTCAGATGATTC 463
    |||||
Db 121 TTGCGGCCCTGAGGGCACCAGCGCTCATCACTTCTCTACTCTGCTCAGATGATTC 180
    |||||
Qy 464 CACCGTTCCACAACGTCATCGACCGCTTGTCTGTAACCTGTCCTCAAGAGCGCTGGTGG 523
    |||||
Db 181 CACCGTTCCACAACGTCATCGACCGCTTGTCTGTAACCTGTCCTCAAGAGCGCTGGTGG 240
    |||||
Qy 524 GCGTTGCGCACTTCTAGTGAGCGGCAAGTACGACCTGCCCCCTGCGCAGATGCCCTGGT 583
    |||||
Db 241 GCGTTGCGCACTTCTAGTGAGCGGCAAGTACGACCTGCCCCCTGCGCAGATGCCCTGGT 300
    |||||
Qy 584 CGGCGCGTTTCTTCTGGCGATCGATCTTTCGACATCGACACATTCATCGGCCCGGAGC 643
    |||||
Db 301 CGGCGCGTTTCTTCTGGCGATCGATCTTTCGACATCGACACATTCATCGGCCCGGAGC 360
    |||||
Qy 644 GCCGCGCTACCTGGAGCAGAAAGCTGGAGCGCTGTGGGAGCAGAAACACCCAGGGTTCGA 703
    |||||
Db 361 GCCGCGCTACCTGGAGCAGAAAGCTGGAGCGCTGTGGGAGCAGAAACACCCAGGGTTCGA 420
    |||||
Qy 704 TCCCTACTAGTCCCTGCTGCGGCCCTTACTACGTTGTGGAATTGGCGCGCTGCCCCAGCG 763
    |||||
Db 421 TCCCTACTAGTCCCTGCTGCGGCCCTTACTACGTTGTGGAATTGGCGCGCTGCCCCAGCG 480
    |||||
Qy 764 TTGGCCACGCGCTTCAGGAGGCGGTGGAGCGCGCCGCTTCTCCGCCCACTTCC 823
    |||||
Db 481 TTGGCCACGCGCTTCAGGAGGCGGTGGAGCGCGCCGCTTCTCCGCCCACTTCC 540
    |||||

```

```

Db      362  TCCCTTACGTGCGGTGGCTGGCGCCCTTACTACTACGTGTGATGTGCNCGCTGCCCCACGC 421
Qy      764  TTGGCCACGCGCTGCACACGAGCGCGTGGAGGGCGCCGCATGTTCCCGCCCA 817
Db      422  TTGGCCACGCGCTGCACACGAGAGCGGTGGAGGGCGCCGCAATGTTCCGCCCC 475

RESULT 4
AV629528          467 bp      mRNA      linear      EST 15-DEC-2000
LOCUS      AV629528 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas
DEFINITION reinhardtii cDNA clone LCL060c10 r 5', mRNA sequence.
ACCESSION AV629528
VERSION   1
KEYWORDS
SOURCE    Chlamydomonas reinhardtii
ORGANISM Chlamydomonas reinhardtii
          Chlamydomonas reinhardtii
          Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;
          Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.
          1 (bases 1 to 467)
REFERENCE 1 (bases 1 to 467)
AUTHORS   Nakamura,Y., Miura,K., Kucho,K., Inoue,Y., Fukuzawa,H., Ohyama,K.,
          Asamizu,E. and Tabata,S.
TITLE      Generation of expressed sequence tags from low-CO2 and high-CO2
          adapted cells of Chlamydomonas reinhardtii
JOURNAL   DNA Res. 7 (5), 305-307 (2000)
PUBMED   11089912
COMMENT    Contact: Erika Asamizu
          The First Laboratory for Plant Gene Research
          Kazusa DNA Research Institute
          Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
          Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
FEATURES
         source
         1..467
            /organism="Chlamydomonas reinhardtii"
            /mol_type="mRNA"
            /db_xref="taxon:3055"
            /strain="C9"
            /clone="LCL060c10 r"
            /notes="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:
            XhoI; The cDNA library was constructed from cells cultured
            in a carbon stress acclimatized condition in which carbon
            dioxide concentration in the bubbling gas was changed from
            5% to 0.04%"

ORIGIN

Query Match      23.8%; Score 463.8; DB 1; Length 467;
Best Local Similarity 99.8%; Pred. No. 4.5e-85;
Matches 465; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      178  GCTGCGCTTTGCTGCCCGCTGGCGGAGCGCTCGAACTCATCTGGGTGACCTGGGTGGT 237
Db      1    GCTGCGAGTTGCTGCCCGCTGGCGGAGCGCTCGAACTCATCTGGGTGACCTGGGTGGT 60

Qy      238  GGCACCTGGGAGAAATGTCGATATGATGGCTGATTATACGACCTGGCGGAAGTTCAAGTCC 297
Db      61    GGCACCTGGGAGAAATGTCGATATGATGGCTGATTATACGACCTGGCGGAAGTTCAAGTCC 120

Qy      298  ATCTAGCTGGTGCACCTGTGCCACTCGCTGTGCGAGGTGGCCCAAGAAGAGGGCAAGGCC 357
Db      121  ATCTAGCTGGTGCACCTGTGCCACTCGCTGTGCGAGGTGGCCCAAGAAGAGGGCAAGGCC 180

Qy      358  AAGGGCTGGAAGAAATGTCAGAGTCTGTGGAGGCGGACGCTTGCCAAATTGGCGCCCTCGAG 417
Db      181  AAGGGCTGGAAGAAATGTCAGAGTCTGTGGAGGCGGACGCTTGCCAAATTGGCGCCCTCGAG 240

Qy      418  GGCACCGCAGCGCTCATACCTTTCTCCTACTCGCTCACGATGATTCACCGTTCCACAAC 477
Db      241  GGCACCGCAGCGCTCATACCTTTCTCCTACTCGCTCACGATGATTCACCGTTCCACAAC 300

Qy      478  GTCATCGACAGGCTTGCTGTCGTAACCTGTGCCAAGAAGCGGCTGGTGGCGCTTGCGGACTTC 537
Db      301  GTCATCGACAGGCTTGCTGTCGTAACCTGTGCCAAGAAGCGGCTGGTGGCGCTTGCGGACTTC 360

```

```

Db      362  TCCCTTACGTGCGGTGCTGCGGCCCCCTTACTACTAGTGTGATGTGCNCGCCTGCCCGCAGC 421
Qy      764  TTGGCCACGCGCTGCACAGAGCGCGTGGAGGGCGGCCCATTTCCCGCCCA 817
Db      422  TTGGCCACGCGCTGCACAGAGCGCGTGGAGGGCGGCCCATTTCCCGCCCA 475

RESULT 4
AV629528          467 bp      mRNA      linear      EST 15-DEC-2000
LOCUS      AV629528 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas
DEFINITION reinhardtii cDNA clone LCL060c10.r 5', mRNA sequence.
ACCESSION AV629528
VERSION    1
KEYWORDS
SOURCE
ORGANISM   Chlamydomonas reinhardtii
            Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;
            Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.
            1 (bases 1 to 467)
REFERENCE   1 (bases 1 to 467)
AUTHORS    Asamizu, Y., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohyama, K.,
            Nakamura, Y. and Tabata, S.
TITLE      Generation of expressed sequence tags from low-CO2 and high-CO2
            adapted cells of Chlamydomonas reinhardtii
JOURNAL    DNA Res. 7 (5), 305-307 (2000)
PUBMED    11089912
COMMENT    Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
FEATURES
         source
         1..467
            /organism="Chlamydomonas reinhardtii"
            /mol_type="mRNA"
            /db_xref="taxon:3055"
            /strain="C9"
            /clone="LCL060c10.r"
            /notes="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:
            XhoI; The cDNA library was constructed from cells cultured
            in a carbon stress acclimatized condition in which carbon
            dioxide concentration in the bubbling gas was changed from
            5% to 0.04%"

ORIGIN

Query Match      23.8%; Score 463.8; DB 1; Length 467;
Best Local Similarity 99.8%; Pred. No. 4.5e-85;
Matches 465; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      178  GCTGCGCTTTGCTGCCCGCTGGCGGAGCGCTCGAACTCATCTGGGTGACCTGGGTGGT 237
Db      1    GCTGCGAGTTGCTGCCCGCTGGCGGAGCGCTCGAACTCATCTGGGTGACCTGGGTGGT 60

Qy      238  GGCACCTGGGAGAAATGTCGATATGATGGCTGATTATACGACCTGGCGGAAGTTCAAGTCC 297
Db      61    GGCACCTGGGAGAAATGTCGATATGATGGCTGATTATACGACCTGGCGGAAGTTCAAGTCC 120

Qy      298  ATCTAGTGGTGCACCTGTGCCACTCGCTGTGCGAGGTGGCCCAAGAAGAGGGCGAAGGCC 357
Db      121  ATCTAGTGGTGCACCTGTGCCACTCGCTGTGCGAGGTGGCCCAAGAAGAGGGCGAAGGCC 180

Qy      358  AAGGGCTGGAAGAAATGTCAGAGTCTGTGGAGGCCGACGCTTGCCAAATTGCGCCCCCTTGAG 417
Db      181  AAGGGCTGGAAGAAATGTCAGAGTCTGTGGAGGCCGAGCGCTTGCCAAATTGCGCCCCCTTGAG 240

Qy      418  GGCACCGCAGCGCTCATACCTTTCTCCTACTCGCTCACGATGATTCACCGTTCCACAAC 477
Db      241  GGCACCGCAGCGCTCATACCTTTCTCCTACTCGCTCACGATGATTCACCGTTCCACAAC 300

Qy      478  GTCATCGACAGGCGTTGCTGTCGTAACCTGTGCCAAGACGGCGCTGGTGGCGCTTGCGGACTTC 537
Db      301  GTCATCGACAGGCGTTGCTGTCGTAACCTGTGCCAAGACGGCGCTGGTGGCGCTTGCGGACTTC 360

```

QY 538 TAGCTGAGCGGCAAGTACGACCTGCCCTCGCCGACAGATGCCCTGCTCGCGCGCTTTCTTC 597  
 Db 361 TAGCTGAGCGGCAAGTACGACCTGCCCTCGCCGACAGATGCCCTGCTCGCGCGCTTTCTTC 420

QY 598 TGGCGATCGCATCTTCGACATCGACAAATTCGACATCGGCCCGCGAGCG 644  
 Db 421 TGGCGATCGCATCTTCGACATCGACAAATTCGACATCGGCCCGCGAGCG 467

RESULT 5  
 AV628087 532 bp mRNA linear EST 15-DEC-2000  
 LOCUS AV628087 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas  
 DEFINITION reinhardtii cDNA clone LCL036e03\_r 5', mRNA sequence.

ACCESSION AV628087  
 VERSION AV628087.1 GI:10790721  
 KEYWORDS EST.  
 SOURCE Chlamydomonas reinhardtii  
 ORGANISM Chlamydomonas reinhardtii  
 Chlamydomonas reinhardtii  
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;  
 Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.

REFERENCE 1 (bases 1 to 532)  
 AUTHORS Asamizu, E., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohyama, K.,  
 Nakamura, Y. and Tabata, S.  
 TITLE Generation of expressed sequence tags from low-CO2 and high-CO2  
 adapted cells of Chlamydomonas reinhardtii  
 JOURNAL DNA Res. 7 (5), 305-307 (2000)  
 PUBMED 11089912  
 COMMENT Contact: Erika Asamizu  
 The First Laboratory for Plant Gene Research  
 Kazusa DNA Research Institute  
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES  
 source  
 1..532  
 Location/Qualifiers  
 /organism="Chlamydomonas reinhardtii"  
 /mol\_type="mRNA"  
 /strain="C9"  
 /db\_xref="taxon:3055"  
 /clone="LCL036e03\_r"  
 /notes="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:  
 XhoI; The cDNA library was constructed from cells cultured  
 in a carbon stress acclimatized condition in which carbon  
 dioxide concentration in the bubbling gas was changed from  
 5% to 0.04%"

ORIGIN  
 Query Match 21.3%; Score 414; DB 1; Length 532;  
 Best Local Similarity 89.8%; Pred. No. 8.5e-75;  
 Matches 478; Conservative 0; Mismatches 0; Indels 54; Gaps 1;

QY 7 TCGGGTCTGTCAGCGCGCGCTCGAGCTACACCAAGAAGAACTTCTCCCTGGAGAGCTC 66  
 Db 1 TCGGGTCTGTCAGCGCGCGCTCGAGCTACACCAAGAAGAACTTCTCCCTGGAGAGCTC 60

QY 67 AAGCTCAGCAGCATGAAGATGACCTGACCGTTCTGGCCATATGTGCTGGCAGCAAG 126  
 Db 61 AAGCTCAGCAGCATGAAGATGACCTGACCGTTCTGGCCATATGTGCTGGCAGCAAG 120

QY 127 AAGGGCGATGATCAGCTGCTCGCTCGAGAGCTTCTACGGGCCCGCAGCGCTGCTC 184  
 Db 121 AAGGGCGATGATCAGCTGCTCGCTCGAGAGCTTCTACGGGCCCGCAGCGCTGCTC 180

QY 185 -----TTCTGCCC 192  
 Db 181 GATGCTTTCCGGTCGCGCTTCTCTGGGTCGAGGCCCATGCTCGCTGCAATTTGCTGCC 240

QY 193 CGCCTGGCCGAGCGCTCGAACCTCATCTGGGTTGACCTGGGTGTGTCACCTGGGGAGAT 252  
 Db 241 CGCCTGGCCGAGCGCTCGAACCTCATCTGGGTTGACCTGGGTGTGTCACCTGGGGAGAT 300

QY 253 GTCGATATGATGGCTGATTACATCGACTGGCGAAGTTCAAGTCCATCTACGTGGTCGAC 312  
 Db 301 GTCGATATGATGGCTGATTACATCGACTGGCGAAGTTCAAGTCCATCTACGTGGTCGAC 360

QY 313 CTGTGCCACTCGCTGTGCGAGGTGGCCAAAGAGAGCGAAGCCAAAGGCTGGAGAAAT 372  
 Db 361 CTGTGCCACTCGCTGTGCGAGGTGGCCAAAGAGAGCGAAGCCAAAGGCTGGAGAAAT 420

QY 373 GTCCAGGTGTGGAGGCCGCGCTTGCCTCAATTTGCCCTCCCTGAGGGCACCGGACGCTC 432  
 Db 421 GTCCAGGTGTGGAGGCCGCGCTTGCCTCAATTTGCCCTCCCTGAGGGCACCGGACGCTC 480

QY 433 ATCACCTTCTCTACTCTCGCTCACGATGATTCCACCGTTCCACAACGTCATCG 484  
 Db 481 ATCACCTTCTCTACTCTCGCTCACGATGATTCCACCGTTCCACAACGTCATCG 532

RESULT 6  
 AV626756 391 bp mRNA linear EST 15-DEC-2000  
 LOCUS AV626756 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas  
 DEFINITION reinhardtii cDNA clone LCL014h09\_r 5', mRNA sequence.

ACCESSION AV626756  
 VERSION AV626756.1 GI:10789036  
 KEYWORDS EST.  
 SOURCE Chlamydomonas reinhardtii  
 ORGANISM Chlamydomonas reinhardtii  
 Chlamydomonas reinhardtii  
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;  
 Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.

REFERENCE 1 (bases 1 to 391)  
 AUTHORS Asamizu, E., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohyama, K.,  
 Nakamura, Y. and Tabata, S.  
 TITLE Generation of expressed sequence tags from low-CO2 and high-CO2  
 adapted cells of Chlamydomonas reinhardtii  
 JOURNAL DNA Res. 7 (5), 305-307 (2000)  
 PUBMED 11089912  
 COMMENT Contact: Erika Asamizu  
 The First Laboratory for Plant Gene Research  
 Kazusa DNA Research Institute  
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES  
 source  
 1..391  
 Location/Qualifiers  
 /organism="Chlamydomonas reinhardtii"  
 /mol\_type="mRNA"  
 /strain="C9"  
 /db\_xref="taxon:3055"  
 /clone="LCL014h09\_r"  
 /notes="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:  
 XhoI; The cDNA library was constructed from cells cultured  
 in a carbon stress acclimatized condition in which carbon  
 dioxide concentration in the bubbling gas was changed from  
 5% to 0.04%"

ORIGIN  
 Query Match 20.1%; Score 391; DB 1; Length 391;  
 Best Local Similarity 100.0%; Pred. No. 4.6e-70;  
 Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 229 CTGGGTGGTGGCACTGGGGAGAAATGTGATATATGATGCTGATTCATCGACCTGGCGAAG 288  
 Db 1 CTGGGTGGTGGCACTGGGGAGAAATGTGATATATGATGCTGATTCATCGACCTGGCGAAG 60

QY 289 TTCAAGTCCATCTACGTGCTGACCTGTCGACCTGTCGACCTGTCGAGGTGGCCAGAAAG 348  
 Db 61 TTCAAGTCCATCTACGTGCTGACCTGTCGACCTGTCGAGGTGGCCAGAAAG 120

QY 349 GCGAAGGCCCAAGGCTGGAAAGAAATGTCCAGGTGTGGAGGCCGACGCTTGCATTTGCG 408  
 Db 121 GCGAAGGCCCAAGGCTGGAAAGAAATGTCCAGGTGTGGAGGCCGACGCTTGCATTTGCG 180

QY 409 CCCCCTGAGGGCACCGCGACGCTCATCACCTTCTCTACTCGCTCACGATGATTCACCG 468



```

181 181 CCCCCCTGAGGCGACCGCGCTCATACCTTCTCTACTCGCTCAGGATATTCACCG 240
QY 469 TTCCACACGTCATCGACGAGCTTGCTCGTACTGTCTCCCAAGACGCGCTGTGGCGGTT 528
Db 241 TTCCACACGTCATCGACGAGCTTGCTCGTACTGTCTCCCAAGACGCGCTGTGGCGGTT 300
QY 529 GCGGACTTCTAGCTGAGCGCAAGTAGCAGCTCCGCTGCGGCGAGATGCCCTGGTCGCGC 588
Db 301 GCGGACTTCTAGCTGAGCGCAAGTAGCAGCTCCGCTGCGGCGAGATGCCCTGGTCGCGC 360
QY 589 CGTTTCTTCTGCGGATCGATCTTCGACATCG 619
Db 361 CGTTTCTTCTGCGGATCGATCTTCGACATCG 391

```

```

RESULT 7
BP086130 379 bp mRNA linear EST 30-JUN-2004
LOCUS BP086130 Chlamydomonas reinhardtii C9 various conditions
DEFINITION Chlamydomonas reinhardtii cDNA clone MX002e03_r 5', mRNA sequence.
ACCESSION BP086130
VERSION BP086130.1 GI:49458217
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;
Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.
REFERENCE 1 (bases 1 to 379)
AUTHORS Asanizu, E., Nakamura, Y., Miura, K., Fukuzawa, H., Fujiwara, S.,
Hirono, M., Iwanoto, K., Matsuda, Y., Minagawa, Y., Shimogawara, K.,
Takahashi, Y. and Tabata, S.
TITLE Establishment of Publicly Available cDNA Material and Information
Resource of Chlamydomonas reinhardtii (Chlorophyta), to Facilitate
Gene Function Analysis
JOURNAL Phycologia (2004) In press
COMMENT Contact: Erika Asanizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asanizu@kazusa.or.jp. URL: http://www.kazusa.or.jp/en/plant/.
FEATURES
source
1..379
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="C9"
/db_xref="taxon:3055"
/clone="MX002e03_r"
/clone_lib="Chlamydomonas reinhardtii C9 various
conditions"
/note="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was made from a mixture of cells
grown under various conditions"

```

```

ORIGIN
Query Match 19.5%; Score 379; DB 3; Length 379;
Best Local Similarity 100.0%; Pred. No. 1.4e-67;
Matches 379; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1449 CCGCATCCCATTTGAACTACATCGCGCGCACCATTGGAGCGGCTGGCGAGACTCGCA 1508
Db 1 CCGCATCCCATTTGAACTACATCGCGCGCACCATTGGAGCGGCTGGCGAGACTCGCA 60
QY 1509 CGTGGCAAGCAGAACTACTTCTACTACAACTGCCTCACCAGCAAGTCTCTCGCGACAA 1568
Db 61 CGTGGCAAGCAGAACTACTTCTACTACAACTGCCTCACCAGCAAGTCTCTCGCGACAA 120
QY 1569 CTGCCCCACTACTCGCGAGCGGCTTCGCCACCCCTCAAGAGTGGCGTGGTGACAA 1628
Db 121 CTGCCCCACTACTCGCGAGCGGCTTCGCCACCCCTCAAGAGTGGCGTGGTGACAA 180
QY 1629 CTTGACCGTCTCCACCAACTTCTTCTATGAGAGCTCAAGCGCGCACCTACACCAAGT 1688

```

```

Db 181 CCGACCGCTCTCCACCAACTTCTTATGAGGAGCTCAAAGCGCGCACTTACACCAAGGT 240
QY 1689 GATTTCTGATGGACACGCTGGACTGGCTGGATATGCCGTGGCCAAAGAGTGGCCGAGTG 1748
Db 241 GATTTCTGATGGACACGCTGGACTGGCTGGATATGCCGTGGCCAAAGAGTGGCCGAGTG 300
QY 1749 CTTGGCCAAAGCAGTTGGCGCGCGGCGCATCTGTCATCTGGCGCTCGGCTCCCTCAGCCC 1808
Db 301 CTTGGCCAAAGCAGTTGGCGCGCGGCGCATCTGTCATCTGGCGCTCGGCTCCCTCAGCCC 360
QY 1809 GCCTTACGCGGAGCTGATC 1827
Db 361 GCCTTACGCGGAGCTGATC 379

```

```

RESULT 8
AV631627 354 bp mRNA linear EST 15-DEC-2000
LOCUS AV631627 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas
DEFINITION reinhardtii cDNA clone LCL097C01_r 5', mRNA sequence.
ACCESSION AV631627
VERSION AV631627.1 GI:10794261
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;
Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.
REFERENCE 1 (bases 1 to 354)
AUTHORS Asanizu, E., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohshima, K.,
Nakamura, Y. and Tabata, S.
TITLE Generation of expressed sequence tags from low-CO2 and high-CO2
adapted cells of Chlamydomonas reinhardtii
JOURNAL DNA Res. 7 (5), 305-307 (2000)
COMMENT Contact: Erika Asanizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asanizu@kazusa.or.jp. URL: http://www.kazusa.or.jp/en/plant/.
FEATURES
source
1..354
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="C9"
/db_xref="taxon:3055"
/clone="LCL097C01_r"
/clone_lib="Chlamydomonas reinhardtii 5% to 0.04% CO2"
/note="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was constructed from cells cultured
in a carbon stress acclimatized condition in which carbon
dioxide concentration in the bubbling gas was changed from
5% to 0.04%"

```

```

ORIGIN
Query Match 18.2%; Score 354; DB 1; Length 354;
Best Local Similarity 100.0%; Pred. No. 1.9e-62;
Matches 354; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 329 GCGAGTGTCGCAAGAGAGCGCAAGCGGCTGGAGATGTCCAGAGTGTGGAGG 388
Db 1 GCGAGTGTCGCAAGAGAGAGCGGCTGGAGATGTCCAGAGTGTGGAGG 60
QY 389 CCGAGCGTTTGCACATTTTGGCGCCCTCGAGGCGACCGCGCATCATCTCTCTACT 448
Db 61 CCGAGCGTTTGCACATTTTGGCGCCCTCGAGGCGACCGCGCATCATCTCTCTACT 120
QY 449 CGCTCAGCATGATTCACCGTTTCCACAGCTCATGACCAAGCGTGTCTGTACTCTGCC 508
Db 121 CGCTCAGCATGATTCACCGTTTCCACAGCTCATGACCAAGCGTGTCTGTACTCTGCC 180
QY 509 AAGAGCGGCTGTGGCGGTTGGCGGCTTCTTCTAGTGGCGGCAAGTACGACCTGCGCCCTGC 568
Db 181 AAGAGCGGCTGTGGCGGTTGGCGGCTTCTTCTAGTGGCGGCAAGTACGACCTGCGCCCTGC 240

```

```

QY 569 GCCAGATGCCCTGTCGCGCCGTTTCTTCTGCGGATCGATCTTTCGACATCGACAACTTG 628
Db 241 GCCAGATGCCCTGTCGCGCCGTTTCTTCTGCGGATCGATCTTTCGACATCGACAACTTG 300

QY 629 ACATCGCGCCCGAGCGCGCGCTTACCTGGAGCAGAGCTGGAGCGCGTGTGGG 682
Db 301 ACATCGCGCCCGAGCGCGCGCTTACCTGGAGCAGAGCTGGAGCGCGTGTGGG 354

RESULT 9
BP093875 352 bp mRNA linear EST 30-JUN-2004
LOCUS BP093875 Chlamydomonas reinhardtii C9 various conditions
DEFINITION Chlamydomonas reinhardtii cDNA clone MXL015a05_r 5', mRNA sequence.
ACCESSION BP093875
VERSION BP093875.1 GI:49465962
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;
Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.
REFERENCE 1 (bases 1 to 352)
AUTHORS Asamizu,E., Iwamoto,K., Matsuda,Y., Minagawa,J., Shimogawara,K.,
Takahashi,Y. and Tabata,S.
TITLE Establishment of Publicly Available cDNA Material and Information
Resource of Chlamydomonas reinhardtii (Chlorophyta), to Facilitate
Gene Function Analysis
JOURNAL Phycologia (2004) In press
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.

FEATURES
source
1..352
Location/Qualifiers
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="C9"
/db_xref="taxon:3055"
/clone="MXL015a05_r"
/clone_lib="Chlamydomonas reinhardtii C9 various
conditions"
/notes="vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was made from a mixture of cells
grown under various conditions"

ORIGIN
Query Match 18.1%; Score 352; DB 3; Length 352;
Best Local Similarity 100.0%; Pred. No. 5e-62;
Matches 352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1257 CGCCAAACGCGCCCAATGGAGGAGCAGCGCGTCTGTGGACAGCAACATGCTCATCCA 1316
Db 1 CGCCAAACGCGCCCAATGGAGGAGCAGCGCGTCTGTGGACAGCAACATGCTCATCCA 60

QY 1317 CTTGCTGAAGAA CGGGCCCAAGCCGCTGGTGTGCTGTTTGTGTAAGTTCTGAGCCCTGGT 1376
Db 61 CTTGCTGAAGAAACGGGCCCCAAGCCGCTGGTGTGCTGTTTGTGTAAGTTCTGAGCCCTGGT 120

QY 1377 GCTCTTCAACAAAGCCGTGCTGTTGTTGCGGCGCGCGTGCAGGAGCAGTACGCGCT 1436
Db 121 GCTCTTCAACAAAGCCGTGCTGTTGTTGCGGCGCGCGTGCAGGAGCAGTACGCGCT 180

QY 1437 GATCAAGCGGACGGCATCCCCATTGAGAACTACATCGCGCACCACATGAGCGCGTGGC 1496
Db 181 GATCAAGCGGACGGCATCCCCATTGAGAACTACATCGCGCACCACATGAGCGCGTGGC 240

QY 1497 GGAGAACTCGACGCTGGCAGCAGAACTACTTCTACTAGAACTGCTTCAACCGGCAAGTT 1556
Db 241 GGAGAACTCGACGCTGGCAGCAGAACTACTTCTACTAGAACTGCTTCAACCGGCAAGTT 300

```

```

QY 1557 CCTGCGCGCAAACTGCCCACTTACCTGCGCGAGCGGCTTGGCCACCTC 1608
Db 301 CCTGCGCGCAAACTGCCCACTTACCTGCGCGAGCGGCTTGGCCACCTC 352

RESULT 10
BP093208 454 bp mRNA linear EST 30-JUN-2004
LOCUS BP093208 Chlamydomonas reinhardtii C9 various conditions
DEFINITION Chlamydomonas reinhardtii cDNA clone MXL005a04_r 5', mRNA sequence.
ACCESSION BP093208
VERSION BP093208.1 GI:49465295
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;
Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.
REFERENCE 1 (bases 1 to 454)
AUTHORS Asamizu,E., Iwamoto,K., Matsuda,Y., Minagawa,J., Shimogawara,K.,
Takahashi,Y. and Tabata,S.
TITLE Establishment of Publicly Available cDNA Material and Information
Resource of Chlamydomonas reinhardtii (Chlorophyta), to Facilitate
Gene Function Analysis
JOURNAL Phycologia (2004) In press
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.

FEATURES
source
1..454
Location/Qualifiers
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="C9"
/db_xref="taxon:3055"
/clone="MXL005a04_r"
/clone_lib="Chlamydomonas reinhardtii C9 various
conditions"
/notes="vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was made from a mixture of cells
grown under various conditions"

ORIGIN
Query Match 17.3%; Score 336; DB 3; Length 454;
Best Local Similarity 86.3%; Pred. No. 1e-58;
Matches 385; Conservative 0; Mismatches 52; Indels 9; Gaps 1;

QY 69 GCTCAGCAGCATGAAGATGACCTGACCGTTCTGCGGCATATGTGTTGCGGCAAGAA 128
Db 9 GTTCGCGCAGCAAGAGGGCGATGATCACGCTGCTCGCTGGAGAGCTTCTACGGGCCCA 68

QY 129 GGGCGATGATCAGCTGCTCGCTTGGAGAGCTTC-----TACGGGCCCAAGGCCG 179
Db 69 GGGCGCTGCCTATGATGCTTTCGGTGTGCGGTTCTCTGGGGTGCAGAGCCATGCTCGC 128

QY 180 TGCCTTTGCTGCCCGCTGCGCGCAGCGCTCGAACCTCATCTGGGTTGACCTGGGTGG 239
Db 129 TGCACTTGTCTGCCCGCTGCGCGCAGCGCTCGAACCTCATCTGGTTCACCTGGGTGG 188

QY 240 CACTGGGAGAAATGTGATATGATGCTGATATCATCGACTGGCGAAGTTCAAGTCCAT 299
Db 189 CACTGGGAGAAATGTGATATGATGCTGATATCATCGACTGGCGAAGTTCAAGTCCAT 248

QY 300 CTAGTGGTGCACCTGTGCCACTGCTGCGAGTGGCCCAAGAGAGGCGAAGGCCAA 359
Db 249 CTAGTGGTGCACCTGTGCCACTGCTGCGAGTGGCCCAAGAGAGGCGAAGGCCAA 308

QY 360 GGGCTGGAAGAAATGTCCAGGTCGTGGAGGCGCGAGCTTGGCAATTTGCGCCCTGAGGG 419
Db 309 GGGCTGGAAGAAATGTCCAGGTCGTGGAGGCGCGAGCTTGGCAATTTGCGCCCTGAGGG 368

QY 420 CACGCGAGCGCTCATACCTTCTCTCTCTGCTCAGATGATTCACCGTTCACAACGT 479

```

Db	369	CACCGCAGCGCTCAACCTTCTCCTACTGCTCACGATGATTCACCGTTCCACAACGT	428
Qy	480	CATCGACCAGCGTTGCTCGTACCTGT	505
Db	429	CATCGACCAGCGTTGCTCGTACCTGT	454
RESULT 11			
AV641303			
LOCUS	AV641303	Chlamydomonas reinhardtii 5% CO2 linear EST 15-DEC-2000	
DEFINITION	CDNA clone HCL031d02_r 5', mRNA sequence.		
ACCESSION	AV641303		
VERSION	AV641303.1	GI:10784631	
KEYWORDS	EST.		
SOURCE	Chlamydomonas reinhardtii		
ORGANISM	Chlamydomonas reinhardtii		
REFERENCE	Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;		
AUTHORS	Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas. 1 (bases 1 to 508)		
TITLE	Asamizu,Y., Miura,K., Kucho,K., Inoue,Y., Fukuzawa,H., Ohyama,K., Nakamura,Y. and Tabata,S.		
JOURNAL	Generation of expressed sequence tags from low-CO2 and high-CO2 adapted cells of Chlamydomonas reinhardtii		
PUBMED	DNA Res. 7 (5), 305-307 (2000)		
COMMENT	11089912 Contact: Erika Asamizu The First Laboratory for Plant Gene Research Kazusa DNA Research Institute Yana 1532-3, Kisarazu, Chiba 292-0812, Japan Email: asamizuka@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/. Location/Qualifiers 1..508 /organism="Chlamydomonas reinhardtii" /mol_type="mRNA" /strain="C9" /db_xref="taxon:3055" /clone="HCL031d02_r" /note="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2: XhoI; The cDNA library was constructed from cells cultured in a medium with bubbling air containing 5% carbon dioxide"		
FEATURES			
source			
ORIGIN			
Query Match	16.4%;	Score 319;	DB 1; Length 508;
Best Local Similarity	87.6%;	Pred. No. 3.3e-55;	
Matches	383;	Conservative 0; Mismatches 0; Indels 54; Gaps 1;	
Qy	1	ATGGGGTTCGGGTCTGCACGCCCGGCTCGCAGCTACACCAAGAAGAACTTCTCCCTGGAG	60
Db	72	ATGGGGTTCGGGTCTGCACGCCCGGCTCGCAGCTACACCAAGAAGAACTTCTCCCTGGAG	131
Qy	61	AGACTCAAGCTCAGCAGCATGAAGAGTAGCCTGACCGTTCTGCGCCATATGTGGTTCGGC	120
Db	132	AGACTCAAAGCTCAGCAGCATGAAGAGTAGCCTGACCGTTCTGCGCCATATGTGGTTCGGC	191
Qy	121	AGCAAGAAGGCGCATGATCACCGTGCTCGCTCGGAGAGCTTCTACGGGCGCCAGGCGGT	180
Db	192	AGCAAGAAGGCGCATGATCACCGTGCTCGCTCGGAGAGCTTCTACGGGCGCCAGGCGGT	251
Qy	181	GCCT-----	186
Db	252	GCCTATGATGCTTTCGGGTTCGGGTTCGCGGCCAATGCTCGCTGCAGTT	311
Qy	187	GCTGCCCGCTTGGCGAGCGCTCGAACCTCATCTGGGTTGACCTCGGTTGGTGGCACTGGG	246
Db	312	GCTGCCCGCTTGGCGAGCGCTCGAACCTCATCTGGTTGACCTGGGTTGGTGGCACTGGG	371
Qy	247	GAGAATGTCGATATGATGGCTGANTATCATCGACTCGCGAAGTTCAAGTCCATCTACGTG	306
Db	372	GAGAATGTCGATATGATGGCTGANTATCATCGACTCGCGAAGTTCAAGTCCATCTACGTG	431

Qy 247 GAGAAATGCGATATGATGGCTGATTACATCGACTGGCGAAGTTCAAGTCCATCTACGTG 306  
Db 384 GAGAAATGCGATATGATGGCTGATTACATCGACTGGCGAAGTTCAAGTCCATCTACGTG 443  
Qy 307 GTGCACTGTGCCACTCGCTGTGCGAGTGCGCCAAAGAGGCAAGGCGCAAGGGCTGG 366  
Db 444 GTGCACTGTGCCACTCGCTGTGCGAGTGCGCCAAAGAGGCAAGGCGCAAGGGCTGG 503  
Qy 367 AAGAAATGT 374  
Db 504 AAGAAATGT 511

RESULT 13  
AV642589  
LOCUS AV642589 478 bp mRNA linear EST 15-DEC-2000  
DEFINITION AV642589 Chlamydomonas reinhardtii 5% CO2 Chlamydomonas reinhardtii  
CDNA clone HCL054c11\_r 5', mRNA sequence.  
ACCESSION AV642589  
VERSION AV642589.1 GI:10785917  
KEYWORDS EST.  
SOURCE Chlamydomonas reinhardtii  
ORGANISM Chlamydomonas reinhardtii  
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;  
Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.  
REFERENCE 1 (bases 1 to 478)  
AUTHORS Nakamura, Y., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohyama, K.,  
Asamizu, E., and Tabata, S.  
TITLE Generation of expressed sequence tags from low-CO2 and high-CO2  
adapted cells of Chlamydomonas reinhardtii  
JOURNAL DNA Res. 7 (5), 305-307 (2000)  
PUBMED 11089912  
COMMENT Contact: Erika Asamizu  
The First Laboratory for Plant Gene Research  
Kazusa DNA Research Institute  
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES  
source  
1..478  
/organism="Chlamydomonas reinhardtii"  
/mol\_type="mRNA"  
/strain="C9"  
/db\_xref="taxon:3055"  
/clone="HCL054c11\_r"  
/note="Vector: pBluescriptII SK-; Site\_1: EcoRI; Site\_2:  
XhoI; The cDNA library was constructed from cells cultured  
in a medium with bubbling air containing 5% carbon  
dioxide"

ORIGIN  
Query Match 14.7%; Score 287; DB 1; Length 478;  
Best Local Similarity 86.7%; Pred. No. 1.3e-48;  
Matches 351; Conservative 0; Mismatches 0; Indels 54; Gaps 1;  
Qy 1 ATGGGGTGGGTGCTGACGGCGCGCTGCGAGCTACACCAAGAGAACTTCTCCCTGGAG 60  
Db 74 ATGGGGTGGGTGCTGACGGCGCGCTGCGAGCTACACCAAGAGAACTTCTCCCTGGAG 133  
Qy 61 AAGCTCAAGCTCAGCAGCATGAAGGATGACCTGACCGTTCTGCGCCATATGTGTTCCGC 120  
Db 134 AAGCTCAAGCTCAGCAGCATGAAGGATGACCTGACCGTTCTGCGCCATATGTGTTCCGC 193  
Qy 121 AGCAAGAGGGCGATGATCAGCTGTGCGCTGAGAGCTTCTACGGGCCCGCCGCGCT 180  
Db 194 AGCAAGAGGGCGATGATCAGCTGTGCGCTGAGAGCTTCTACGGGCCCGCCGCGCT 253  
Qy 181 GCCT-----TT 186  
Db 254 GCCTATGATGTTTCCGGTCCGGTTCTCTGCGGTGCGAGGCCCATGCTCGTGCAGTT 313  
Qy 187 GCTGCCCGCTGGCGGCGCTCGAAGCTCATCTGCGGTGACCTGGGTGGTGGCACTGGG 246

Db 314 GCTGCCCGCTGGCGGCGCTCGAAGCTCATCTGGGTGACCTGGTGGTGGCACTGGG 373  
Qy 247 GAGAAATGCGATATGATGGCTGATTACATCGACTGGCGAAGTTCAAGTCCATCTACGTG 306  
Db 374 GAGAAATGCGATATGATGGCTGATTACATCGACTGGCGAAGTTCAAGTCCATCTACGTG 433  
Qy 307 GTGCACTGTGCCACTCGCTGTGCGAGTGCGCCAAAGAGGCG 351  
Db 434 GTGCACTGTGCCACTCGCTGTGCGAGTGCGCCAAAGAGGCG 478

RESULT 14  
AV643824  
LOCUS AV643824 480 bp mRNA linear EST 15-DEC-2000  
DEFINITION AV643824 Chlamydomonas reinhardtii 5% CO2 Chlamydomonas reinhardtii  
CDNA clone HCL076g02\_r 5', mRNA sequence.  
ACCESSION AV643824  
VERSION AV643824.1 GI:10787152  
KEYWORDS EST.  
SOURCE Chlamydomonas reinhardtii  
ORGANISM Chlamydomonas reinhardtii  
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;  
Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.  
REFERENCE 1 (bases 1 to 480)  
AUTHORS Nakamura, Y., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohyama, K.,  
Asamizu, E., and Tabata, S.  
TITLE Generation of expressed sequence tags from low-CO2 and high-CO2  
adapted cells of Chlamydomonas reinhardtii  
JOURNAL DNA Res. 7 (5), 305-307 (2000)  
PUBMED 11089912  
COMMENT Contact: Erika Asamizu  
The First Laboratory for Plant Gene Research  
Kazusa DNA Research Institute  
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES  
source  
1..480  
/organism="Chlamydomonas reinhardtii"  
/mol\_type="mRNA"  
/strain="C9"  
/db\_xref="taxon:3055"  
/clone="HCL076g02\_r"  
/note="Vector: pBluescriptII SK-; Site\_1: EcoRI; Site\_2:  
XhoI; The cDNA library was constructed from cells cultured  
in a medium with bubbling air containing 5% carbon  
dioxide"

ORIGIN  
Query Match 14.7%; Score 286; DB 1; Length 480;  
Best Local Similarity 86.6%; Pred. No. 2.1e-48;  
Matches 350; Conservative 0; Mismatches 0; Indels 54; Gaps 1;  
Qy 1 ATGGGGTGGGTGCTGACGGCGCGCTGCGAGCTACACCAAGAGAACTTCTCCCTGGAG 60  
Db 77 ATGGGGTGGGTGCTGACGGCGCGCTGCGAGCTACACCAAGAGAACTTCTCCCTGGAG 136  
Qy 61 AAGCTCAAGCTCAGCAGCATGAAGGATGACCTGACCGTTCTGCGCCATATGTGTTCCGC 120  
Db 137 AAGCTCAAGCTCAGCAGCATGAAGGATGACCTGACCGTTCTGCGCCATATGTGTTCCGC 196  
Qy 121 AGCAAGAGGGCGATGATCAGCTGTGCGCTGAGAGCTTCTACGGGCCCGCCGCGCT 180  
Db 197 AGCAAGAGGGCGATGATCAGCTGTGCGCTGAGAGCTTCTACGGGCCCGCCGCGCT 256  
Qy 181 GCCT-----TT 186  
Db 257 GCCTATGATGTTTCCGGTCCGGTTCTCTGCGGTGCGAGGCCCATGCTCGTGCAGTT 316  
Qy 187 GCTGCCCGCTGGCGGCGCTCGAAGCTCATCTGCGGTGACCTGGGTGGTGGCACTGGG 246  
Db 317 GCTGCCCGCTGGCGGCGCTCGAAGCTCATCTGCGGTGACCTGGGTGGTGGCACTGGG 376

QY 247 GAGATGTCGATATGCTGCTGATTACATCGACTGGCGAAGTTCAAGTCCATCTACGTG 306  
|||||  
Db 377 GAGATGTCGATATGCTGCTGATTACATCGACTGGCGAAGTTCAAGTCCATCTACGTG 436  
|||||  
QY 307 GTCGACCTGTGCCACTCGCTGTGCGAGTGGCCAGGAAGAGGC 350  
|||||  
Db 437 GTCGACCTGTGCCACTCGCTGTGCGAGTGGCCAGGAAGAGGC 480  
|||||

RESULT 15  
AV628989  
LOCUS  
DEFINITION AV628989 Chlamydomonas reinhardtii 5' to 0.04% CO2 Chlamydomonas  
reinhardtii cDNA clone LCL050a06\_r 5', mRNA sequence.  
ACCESSION AV628989  
VERSION AV628989  
KEYWORDS EST.  
SOURCE Chlamydomonas reinhardtii  
ORGANISM Chlamydomonas reinhardtii  
Chlamydomonas reinhardtii  
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;  
Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.  
REFERENCE  
AUTHORS Asamizu, E., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohyama, K.,  
Nakamura, Y. and Tabata, S.  
TITLE Generation of expressed sequence tags from low-CO2 and high-CO2  
adapted cells of Chlamydomonas reinhardtii  
JOURNAL DNA Res. 7 (5), 305-307 (2000)  
PUBMED 11089912  
COMMENT Contact: Erika Asamizu  
The First Laboratory for Plant Gene Research  
Kazusa DNA Research Institute  
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES  
source  
1..466  
/organism="Chlamydomonas reinhardtii"  
/mol\_type="mRNA"  
/strain="C9"  
/db\_xref="taxon:3055"  
/clone="LCL050a06\_r"  
/note="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:  
XhoI; The cDNA library was constructed from cells cultured  
in a carbon stress acclimatized condition in which carbon  
dioxide concentration in the bubbling gas was changed from  
5% to 0.04%"

ORIGIN  
Query Match 13.7%; Score 267; DB 1; Length 466;  
Best Local Similarity 86.0%; Pred. No. 1.7e-44;  
Matches 331; Conservative 0; Mismatches 0; Indels 54; Gaps 1;  
QY 1 ATGGGGTGGGTGCTGACGGCGCGCTGCGAGCTACACCAAGAACTTCTCCCTGGAG 60  
|||||  
Db 82 ATGGGGTGGGTGCTGACGGCGCGCTGCGAGCTACACCAAGAACTTCTCCCTGGAG 141  
|||||  
QY 61 AGCTCAAGCTCAGCAGCATGAAGGATGACCTGACCGTTCTGCGCCATATGTTGCGC 120  
|||||  
Db 142 AGCTCAAGCTCAGCAGCATGAAGGATGACCTGACCGTTCTGCGCCATATGTTGCGC 201  
|||||  
QY 121 AGCAAGAAGCGGATGATACGCTGCTGCGCTGAGAGCTTCTACGGGCCCAAGCGCT 180  
|||||  
Db 202 AGCAAGAAGCGGATGATACGCTGCTGCGCTGAGAGCTTCTACGGGCCCAAGCGCT 261  
|||||  
QY 181 GCCT-----TT 186  
|||||  
Db 262 GCCTATGATGCTTTCGGTTCGGTTCCTCTGGGGTGGCAGGCCCATGCTCGCTGCAGTT 321  
|||||  
QY 187 GCTGCCCGCTGCGCGAGCGCTCGAACTCATCTGGGTTGACCTGGGTGGTGGCACTGGG 246  
|||||  
Db 322 GCTGCCCGCTGCGCGAGCGCTCGAACTCATCTGGGTTGACCTGGGTGGTGGCACTGGG 381  
|||||  
QY 247 GAGATGTCGATATGATGCTGCTGATTACATCGACTGGCGAAGTTCAAGTCCATCTACGTG 306  
|||||

Db 382 GAGATGTCGATATGATGCTGATTACATCGACTGGCGAAGTTCAAGTCCATCTACGTG 441  
|||||  
QY 307 GTCGACCTGTGCCACTCGCTGTGCGG 331  
|||||  
Db 442 GTCGACCTGTGCCACTCGCTGTGCGG 466  
|||||

Search completed: March 14, 2006, 01:11:07  
Job time : 7968 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2006, 01:55:41 ; Search time 131 Seconds  
(without alignments)  
3489.942 Million cell updates/sec

Title: US-10-620-914-45  
Perfect score: 3463  
Sequence: 1 MGSGRDRPASVTKKNFSL.....RVNYSFYMARKGAKDN 648

Scoring table:  
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 05.80.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3441	99.4	666	Q6DN05	CHLRE
2	953	27.5	835	Q4IPW6	GIBZE
3	924	26.7	944	Q4PF55	USTLAGO
4	910	26.3	741	Q6CE83	YARLI
5	891	25.7	752	Q59TS3	CANAL
6	890.5	25.7	790	Q6CJ18	KLUIA
7	881.5	25.5	832	Q52GF6	MAGR
8	881	25.4	820	Q55J00	CRYNE
9	864.5	25.0	763	Q6BK65	DEBHA
10	850.5	24.6	908	Q7SGV5	NEUCR
11	845.5	24.4	797	Q5KXJ9	CRYNE
12	806.5	23.3	831	Q4X175	ASPU
13	752	21.7	790	Q5B016	EMENI
14	458	13.2	415	Q7UYV8	RHOBA
15	409.5	11.8	251	Q7UYV9	RHOBA
16	380.5	11.0	407	Q6NBV1	RHOBA
17	379.5	11.0	416	Q93TQ1	RHOSH
18	326.5	9.4	416	Q8UDK6	AGRY5
19	321	9.3	432	Q98K98	RHILO
20	292.5	8.4	416	Q92NK6	RHIME
21	161	4.6	367	Q6MIJ9	BDEBA
22	159	4.6	377	Q5YZV2	NOCPA
23	123	3.6	2283	Q80XH7	MOUSE
24	122	3.5	2283	Q92NK5	RHIME
25	120	3.5	221	Q92NK5	RHIME
26	120	3.5	451	Q92NK5	RHIME
27	120	3.5	716	Q8CD66	MOUSE
28	119	3.4	983	Q59EA9	HUMAN
29	119	3.4	2243	Q9UNF3	HUMAN
30	119	3.4	2286	Q9UNF3	HUMAN
31	119	3.4	2286	Q9UNF3	HUMAN

32	119	3.4	2297	2	Q9Y5S5	HUMAN
33	114.5	3.3	1128	2	Q5F3Q4	CHICK
34	114.5	3.3	1920	2	Q6BNV7	DEBHA
35	112.5	3.2	672	2	Q8VUQ5	PSEPU
36	112.5	3.2	672	2	Q88PN6	PSEPK
37	111	3.2	477	2	Q6NCK7	RHOBA
38	110.5	3.2	782	2	Q4JCH8	SULAC
39	109.5	3.2	236	2	Q7NYSY	CHRVO
40	109.5	3.2	853	2	Q92MT9	RHIME
41	109	3.1	198	2	Q65GZ7	BACID
42	108.5	3.1	222	2	Q8TN85	METAC
43	108.5	3.1	245	2	Q6NEV2	RHOBA
44	108	3.1	255	2	Q6NTG7	HUMAN
45	108	3.1	676	2	Q64QR3	BACPR

ALIGNMENTS

RESULT 1  
Q6DN05\_CHLRE PRELIMINARY; PRT; 666 AA.  
ID Q6DN05; AC Q6DN05;  
DT 25-OCT-2004 (T-EMBLrel. 28, Created)  
DT 25-OCT-2004 (T-EMBLrel. 28, Last sequence update)  
DT 25-OCT-2004 (T-EMBLrel. 28, Last annotation update)  
DE Betaine lipid synthase.  
GN Names=BTA1;  
OS Chlamydomonas reinhardtii.  
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;  
OC Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.  
OX NCBI\_TaxID=3055;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=CC125;  
RX PubMed=15701786; DOI=10.1128/EC.4.2.242-252.2005;  
RA Riekhof W.R., Sears B.B., Benning C.;  
RT "Annotation of Genes involved in Glycerolipid Biosynthesis in  
Chlamydomonas reinhardtii: Discovery of the Betaine Lipid Synthase  
BTA1Cr.";  
RL Eukaryot. Cell 4:242-252(2005).  
DR EMBL: AY656806; AN77342.1; -; mRNA.  
DR GO: GO:0008757; F:S-adenosylmethionine-dependent methyltransf. ...; IEA.  
DR GO: GO:0016740; F:transferase activity; IEA.  
DR InterPro: IPR002114; HPR SerP S.  
DR InterPro: IPR000051; SAM\_bind.  
DR PROSITE: PS00589; PTS HPR SER; UNKNOWN\_1.  
KW Methyltransferase; Transferase.  
SQ SEQUENCE 666 AA; 75780 MW; 4B6AF5718682284 CRC64;

Query Match	99.4%;	Score	3441;	DB	2;	Length	666;		
Best Local Similarity	97.1%;	Pred.	No. 6.6e-273;						
Matches	647;	Conservative	1;	Mismatches	0;	Indels	18;	Gaps	1;
Qy	1	MGSGRDRPASVTKKNFSL	KLSSMKDDLTVLRHMFSGSKGDDHAARLESFYGPQAA	60					
Db	1	MGSGRDRPASVTKKNFSL	KLSSMKDDLTVLRHMFSGSKGDDHAARLESFYGPQAA	60					
Qy	61	AF-----	AAALBAERSNLIWDLGGGTGENVDMADYIDLAKFKSIYV	102					
Db	61	AYDAFASRFLWGRRPMLAAVAARBAERSNLIWDLGGGTGENVDMADYIDLAKFKSIYV	120						
Qy	103	VDLCHSLCEVAKKAKAGKNVQVVEADACQAPPEGTATLITFSYSLTWIPFFHNVID	162						
Db	121	VDLCHSLCEVAKKAKAGKNVQVVEADACQAPPEGTATLITFSYSLTWIPFFHNVID	180						
Qy	163	QACSYLSQDLGVGVADFYVSGKYDPLRQMPNSRRFFWRSIFDIDNIDIGPERRAYLEQK	222						
Db	181	QACSYLSQDLGVGVADFYVSGKYDPLRQMPNSRRFFWRSIFDIDNIDIGPERRAYLEQK	240						
Qy	223	LERVWEQNTQGSIPYVWLRAPYVYVWIGRLPSVGHALHEERVERPMPFPPTFLYTQSWED	282						
Db	241	LERVWEQNTQGSIPYVWLRAPYVYVWIGRLPSVGHALHEERVERPMPFPPTFLYTQSWED	300						

QY 283 PEPDMEVMEINPKDVTLTGSGCNALNLVQAGGVVSDCNPAQSALLEKKVAIQOL 342  
 DB 301 PEPDMEVMEINPKDVTLTGSGCNALNLVQAGGVVSDCNPAQSALLEKKVAIQOL 360  
 QY 343 EFEDVWOLFGEVHPRIEELYEKKLAPFLSQTSHNFWSKELMTFOHGLYYQGGMGKLCW 402  
 DB 361 EFEDVWOLFGEVHPRIEELYEKKLAPFLSQTSHNFWSKELMTFOHGLYYQGGMGKLCW 420  
 QY 403 LQCLAVVLGLGKTVKRLANAPTHMEEQRLNDSNMLHFVXNGPKPLVWLFVSVLFLN 462  
 DB 421 LQCLAVVLGLGKTVKRLANAPTHMEEQRLNDSNMLHFVXNGPKPLVWLFVSVLFLN 480  
 QY 463 KAVLWFGGGVPGKQYALIKADGPIENYIARTWDGVAENSHVRKONYFYNYCLTGKFLRD 522  
 DB 481 KAVLWFGGGVPGKQYALIKADGPIENYIARTWDGVAENSHVRKONYFYNYCLTGKFLRD 540  
 QY 523 NCPTYLREAAAFATLKSGVVDNLTVSTNFFMEELKARTYTKVILMDHVDLMDPVAEELAE 582  
 DB 541 NCPTYLREAAAFATLKSGVVDNLTVSTNFFMEELKARTYTKVILMDHVDLMDPVAEELAE 600  
 QY 583 CLAKQVAPGGIVWRSASLSPVAELIQKAGFVRCIRRTATQGMORVNNYSSFYVARRK 642  
 DB 601 CLAKQVAPGGIVWRSASLSPVAELIQKAGFVRCIRRTATQGMORVNNYSSFYVARRK 660  
 QY 643 GAKKON 648  
 DB 661 GAKKON 666

## RESULT 2

Q4IPW6\_GIBZE PRELIMINARY; PRT; 835 AA.  
 ID Q4IPW6\_GIBZE PRELIMINARY; PRT; 835 AA.  
 AC Q4IPW6;  
 DT 13-SEP-2005 (TrEMBLrel. 31, Created)  
 DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)  
 DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)  
 DE Hypothetical protein.  
 GN ORFNames=FG00742.1;  
 OS Gibberella zeae PH-1.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Eupocreomycetidae; Hypocreales; Nectriaceae; Gibberella.  
 OX NCBI\_TaxID=229533;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=PH-1;  
 RA Birren B., Nusbaum C., Abouelleil A., Allen N., Anderson S.,  
 RA Arachchi H.M., Barna N., Bastien V., Bloom T., Boguslavskiy L.,  
 RA Boukhalter B., Butler J., Calvo S.E., Camarata J., Chang J.,  
 RA Choepel Y., Collymore A., Cook A., Cooke P., Corum B., DeArellano K.,  
 RA Diaz J.S., Dodge S., Dooley K., Dorris L., Elkins T., Engels R.,  
 RA Erickson J., Faro S., Ferreira P., Fitzgerald M., Gage D., Galagan J.,  
 RA Gardyna S., Gnerre S., Graham L., Grand-Pierre N., Hafez N.,  
 RA Hagopian D., Hagos B., Hall J., Horton L., Hulme W., Iliev I.,  
 RA Jaffe D., Johnson R., Jones C., Kamal M., Kamat A., Karatas A.,  
 RA Kells C., Landers T., Levine R., Lindblad-Toh K., Liu G., Lui A.,  
 RA Ma L.-J., Mabbitt R., Maclean C., Macdonald P., Major J., Manning J.,  
 RA Matthews C., Mauceli E., McCarthy M., Meldrim J., Meneus L.,  
 RA Mihova T., Mienga V., Murphy T., Naylor J., Nguyen C., Nicol R.,  
 RA Nielsen C.B., Norbu C., O'Connor T., O'Donnell P., O'Neill D.,  
 RA Oliver J., Peterson K., Phunkhang P., Pierre N., Purcell S.,  
 RA Rachupka A., Ranasingh U., Raymond C., Retta R., Rise C., Rogov P.,  
 RA Roman J., Schauer S., Schupbach R., Seaman S., Severy P., Smirnov S.,  
 RA Smith C., Spencer S., Stange-Thomann N., Stojanovic N., Stubbs M.,  
 RA Talamas J., Tesfaye S., Theodore J., Topham K., Travers M.,  
 RA Vassiliev H., Venkataraman V.S., Viel R., Vo A., Wang S., Wilson B.,  
 RA Wu X., Wyman D., Young G., Zainoun J., Zembek L., Zimmer A., Zody M.,  
 RA Lander E.  
 RT "Fusarium graminearum genome sequence."  
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.  
 CC -!- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.

DR EMBL; AACM01000035; EAA70688.1; -; Genomic\_DNA.  
 KW Hypothetical protein.  
 SQ SEQUENCE 835 AA; 93926 MW; 8192222A79ABC086E CRC64;  
 Query Match 27.5%; Score 953; DB 2; Length 835;  
 Best Local Similarity 32.7%; Pred. No. 8,7e-69;  
 Matches 254; Conservative 103; Mismatches 216; Indels 204; Gaps 25;  
 QY 40 GSKKGDDHAALESFYGPQAAAF-----AARLA-----ERSNLIV 75  
 DB 64 GSKGQQAQDA-LESFYKQAGADATRKVLRLGREDMLALVAAQMAKLDKNKTKRVV 122  
 QY 76 DLGGGTGENVDMMADYIDLAK-FKSIYVVDLCHSLCEVAKKAKAKAGKXNVQVVEADACQ 134  
 DB 123 DIGGGTGFNIEAMGAFVDVPTFESSVYLVDSPSLCEVAKKRPDLRLGKXNVKVCEDARK 182  
 QY 135 F-----APPE-----GTALITFSYSLTWIPFPHNVIDQACS 167  
 DB 183 FRLEDYESGMPSPKSIIPRSPALSYFDKPRDFGADLITMSYSLMIPDYYSVIDSVTSL 242  
 QY 168 LSQDGLGVGADFYVSGKYDLPL-----RQMPWRFRFFWRSIFDIDNIDIGPERRAY 218  
 DB 243 LSPQGGIMGVDFVQNKVDFAFNYTGGVLDRHVNFSLRSFWRSWFDLDRVGLSPRRDY 302  
 QY 219 LEQKLERVWEQNTQ-----GSIPYVWLRAPYVWIG--RLPSYVGHALHEEVER-- 266  
 DB 303 LEYKFGTVLNVTRNKGGLGAI-----PYIWLGHCKKPPSSSLPHEIVERIDALVT 354  
 QY 267 --PPMPF-----PTFLY----- 276  
 DB 355 ESPLYPANHGDAITRAISAPERSKAFLTAVSNLSSNLPPLPSFFYQNHWHRIYYDEQ 414  
 QY 277 ----TQ-----SWDEPEPMVEWNEPKDVTLTGSGCNALNLVQAGGVVSD 323  
 DB 415 LPKHTQFKDEYIYAFWEDTRVDERILKLGADDDKVLAITSGDNILSYLLQSPARVHVD 474  
 QY 324 CNPAQSALLEKKVAIQOLEFEFVWOLFGEVHPRIEELYEKKLAPFLSQTSHNFWSKRL 383  
 DB 475 LNPTQNHLLLEKAAASYALTALPYGDFWKIFGDKGHPRELLITKLSPHLSGRAFYWLKNV 534  
 QY 384 WYFQ-----HGLYYQGGMGKLCWVLOCLAVVLGLGKTVKRLANAPTHMEEQRLNDSNMLIH 439  
 DB 535 HVPQSSGYGLYDTGGSRHAIRVFRWIARIFGLQKAVKQLLQAKTLNEQREIWRRI--- 591  
 QY 440 FVXNGPKPLVWLFVKV-SLVLENKAVLWFGGVPGKQYALIKAD----- 483  
 DB 592 -----RPA--LUSKLVCLNVQSESLWALGVKQNLAMIEDHANSDLVKGPKPAK 643  
 QY 484 ---GPIENYIARTMDGVAENSHVRKONYFYNYCLTGKFLRDNCPTYLREAAAFATL-KSG 539  
 DB 644 NTRSHALWHYVWNTLPVAETHIAADNPYYVYVMDGKFPKCHPDYLSPRAHAKLSRPN 703  
 QY 540 VDNLTVSTNFFMEELKAR-----TYTKVILMDHVDLMDPVAEELAE--LAKQVAPGGI 593  
 DB 704 ALDGLAIHTD-ELEEVIARTITPGTLTIVAVVMDMSMDWFDTGSRAAAQITKLNALAMGGR 762  
 QY 594 VIVRSASLSPVAELIQKAGF-----DVRCIRRTATQGMORVNNYSSFYVARR 641  
 DB 763 VLLRSSALTWPYIKBPEGHGFTPKRHGARIDGACI-----DRVNNYASCIWCTK 811  
 RESULT 3  
 Q4IPF5\_USTMA PRELIMINARY; PRT; 944 AA.  
 ID Q4IPF5\_USTMA PRELIMINARY; PRT; 944 AA.  
 AC Q4IPF5;  
 DT 13-SEP-2005 (TrEMBLrel. 31, Created)  
 DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)  
 DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)  
 DE Hypothetical protein.  
 GN ORFNames=UM01159.1;  
 OS Ustilago maydis 521.  
 OC Eukaryota; Fungi; Basidiomycetes; Ustilaginomycetes;  
 OC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.

OX NCBI\_TaxID=237631;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCES.  
 RC STRAIN=521;  
 RA Birren B., Nusbaum C., Abebe A., Abouelleil A., Adekoya E.,  
 RA Bit-zahra M., Allen T., An P., Anderson M., Anderson S.,  
 RA Arachi H., Ambruster J., Bachanteang P., Baldwin J., Barry A.,  
 RA Bayul T., Blithsteyn B., Bloom T., Blye J., Boguslavskiy L.,  
 RA Borowsky M., Boukhalter B., Brunache A., Butler J., Calixte N.,  
 RA Calvo S., Camarata J., Campo K., Chang J., Cheshatsang Y., Citroen M.,  
 RA Collimore A., Considine T., Cook A., Cooke P., Corum B., Cuomo C.,  
 RA David R., Dawoe T., Degray S., Dodge S., Dooley K., Dorje P.,  
 RA Dorjee K., Dorris L., Duffey N., Dupes A., Elkins T., Engels R.,  
 RA Erickson J., Farina A., Fago S., Ferreira P., Fischer H.,  
 RA Fitzgerald M., Foley K., Gage D., Galagan J., Gearin G., Gnerre S.,  
 RA Gnirke A., Coyette A., Graham J., Grandbois E., Gyaltsen K., Hafez N.,  
 RA Hagopian D., Hagos B., Hall J., Hatcher B., Heller A., Higgins H.,  
 RA Honan T., Horn A., Houde N., Hughes L., Hulme W., Husby E., Iliev I.,  
 RA Jaffe D., Jones C., Kamal M., Kamat A., Kamyselis M., Karlsson E.,  
 RA Kells C., Kieu A., Kienner P., Kodira C., Kulbokas E., Labutti K.,  
 RA Lama D., Landers T., Leger J., Levine S., Lewis D., Lewis T.,  
 RA Lindblad-toh K., Liu X., Lokitsang T., Lokitsang Y., Lucien O.,  
 RA Lui A., Ma L.J., Mabbitt R., Macdonald J., Maclean C., Major J.,  
 RA Manning J., Marabella R., Maru K., Matthews C., Mauceli E.,  
 RA McCarthy M., McDonough S., Mcghee T., Meldrim J., Meneus L.,  
 RA Mesirov J., Mihaliev A., Mihova T., Mikkelsen T., Mlenga V., Moru K.,  
 RA Mozes J., Mulraev L., Munson G., Naylor J., News C., Nguyen C.,  
 RA Nguyen N., Nguyen T., Nicol R., Nielsen C., Nizzari M., Norbu C.,  
 RA Norbu N., O'donnell P., Okoawo O., O'leary S., Omotosho B.,  
 RA O'Neill K., Omsan S., Parker S., Perrin D., Phunkhang P., Pignani B.,  
 RA Purcell S., Rachupka T., Ramasany U., Rameau R., Ray V., Raymond C.,  
 RA Retta T., Richardson S., Rise C., Rodriguez J., Rogers J., Rogov P.,  
 RA Rutman M., Schuback R., Seaman C., Settipalli S., Sharpe T.,  
 RA Sheridan J., Sherpa N., Shi J., Smirnov S., Smith C., Sougnez C.,  
 RA Spencer B., Stalker J., Stange-thomann N., Stavropoulos S.,  
 RA Stetson K., Stone C., Stone S., Stubbs N., Talamas J., Tchuanga P.,  
 RA Tenzon P., Tesfaye S., Theodore J., Thoulutsang Y., Topham K.,  
 RA Towey S., Teamlia T., Tsomo N., Vallee D., Vassiliev H.,  
 RA Venkataraman V., Vinson J., Vo A., Wade C., Wang S., Wangchuk T.,  
 RA Wangdi T., Whittaker C., Wilkinson J., Wu Y., Wyman D., Yadav S.,  
 RA Yang S., Yang X., Yeager S., Yee E., Young G., Zainoun J., Zembeck L.,  
 RA Zimmer A., Zody M., Lander E.;  
 RT The genome sequence of *Ustilago maydis*.  
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.  
 CC -!- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 DR EMBL; AACP0100041; EAK81682.1; -; Genomic\_DNA.  
 DR InterPro; IPR000051; SAM\_bd.  
 KW Hypothetical protein; Methyltransferase; Transferase.  
 SQ SEQUENCE 944 AA; 106525 MW; B7C3A32D89132B0D CRC64;  
  
 Query Match 26.7%; Score 924; DB 2; Length 944;  
 Best Local Similarity 30.0%; Pred. No. 2.4e-66;  
 Matches 234; Conservative 113; Mismatches 237; Indels 196; Gaps 22;  
  
 QY 43 KGDHARLESFYGPQAAAF-----AARL--ABRSN-----LIWVDLG 78  
 DB 179 KTAQAEKLDKFFQONQANVYDSTRGLKGRRTTMLKCAQLDQASNPGKPLVWDIG 238  
  
 QY 79 GGTGENVDMAVDYIDIAKPKSIYVVDLCHSLCBVAKKAKAKGKWKVQVVEADACQFAP 138  
 DB 239 GGTGWNIEQWQFPFIDQLSQVYLIDCBPLQVARKRAAKGFKVQVLCQDASQFNMP 298  
  
 QY 139 EGTAT-----LITFSLSMTIPPHNVQDQACVLSQ-DGLGVADVFGYSGKDLPL--- 189  
 DB 299 -GLAAGQKVDLFTCSYSIMIPFYAVLDKINDLDPVTGVGVDFYVSGSG-FLAKS 356  
  
 QY 190 -----ROMPSRRFFWRSIFDIDNIDIGPERRAYLEOKLRVWEQNTQGS--IPYVP 239  
 DB 357 PMTGGTRQCGLWLRWFSWMSFPHIELHPARDYLEHKEFTIKCYNGRNNFIPIF- 415  
 QY 240 WLRAPYVWIG--RLPSVGHALHEERVER-----PPMPPP----- 272

Db 416 -VRIPYIWLGVSRERDRTTKAQAPEVSGNRVVVPSPPELAYMNGEAHTHTSIAAT 474  
 QY 273 ----- 272  
 Db 475 ATATATATGIDADEDASSRPEPARRLLRRVSEATTESDADSDRPLKLELGHPPPLSSPH 534  
 QY 273 -----TFLYTQSWDEPDPEMBVMEINPKDVTVLTLTSGGCAINLL 312  
 Db 535 YQKRWLPFVDFNEFSDMFTWYGTWEDPYVDMQHLDLGKDDSLCITSAGDNALHYA 594  
 QY 313 VQG-AGQVSVDCNPAQSALELKKVAIQOLEPFDVQLEFEGVHPRIEELYEKCLAPFL 371  
 Db 595 VAKPRRIHADVDMNPPCGHLLLEKLACIASLSYDEMWMQFGEGRINFRELLDSKISPYL 654  
 QY 372 SQTSHNFWSKRLWYFQHGLYQCGMKLQWLQCLAVLGLKTVKRLANAPTMEEORRL 431  
 Db 655 SSHAYQFWRLNTRAFDKAFYPRGYSCHALRLAKAFSVTVGRVWEKMTCTANSVEQEV 714  
 QY 432 WDSNMLIHFWKNGPKPLVWLFVKFVSLVFNKAVLWFGGVPKGVKQYALIKADGPIENYI 491  
 Db 715 WDKKLRSTLIN--KPLIRLF-----LSNPAFLWNLALGVPNMNQIIFLNEGVSABQFA 764  
 QY 492 ARTMDGVAENSHVRKQNYFYNNCLTKGFLRDNCPYILREAAFPATIKSGV---VDNLTVS 547  
 Db 765 IDTLDSIPRSRLTKNDNYHQLCLHKYTKQSCPLYLKPDPGFAALKQALQDGLDSPLRH 824  
 QY 548 TNPFFBELKA---RTYTKVILMDHVDWLD-MPVA-----NELAE 582  
 Db 825 TDSIVNVLGFEDGALTRATWDMDFDPVPASRPAPTIKQARRDSKSVSLDREICE 884  
 QY 583 CIAQVAPGGIVWIRASISPPYAEILIKAGFDVRCIR-RATQGYMDRVNMYSSFYMMARR 641  
 Db 885 -LSRVIRKGGAVFYRSAKKPWYHNRHFRKMGFSVQVPHIRETAKPIDNVNMYASFYKATR 943  
  
 RESULT 4  
 ID Q6C6E3 YARLI PRELIMINARY; PRT; 741 AA.  
 AC Q6C6E3;  
 DT 25-OCT-2004 (TrEMBLrel. 28, Created)  
 DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)  
 DE Similar to DEHA0P25982g Debaryomyces hansenii IPF 6093.1.  
 GN OrderedLocusNames=YALIOE10197g;  
 OS Yarrowia lipolytica (Candida lipolytica).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Dipodascaceae; Yarrowia.  
 OX NCBI\_TaxID=4952;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RC STRAIN=CLIB 122 / E 150;  
 RX PubMed=15229592; DOI=10.1038/nature02579;  
 RA Dujon B., Sherman D., Fischer G., Durand P., Casaregola S.,  
 RA Lafontaine I., de Montigny J., Marck C., Neuvéglise C., Talla E.,  
 RA Goffard N., Frangeul L., Aigle M., Anchoard V., Babour A., Barbe V.,  
 RA Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykasten C.,  
 RA Boisarame A., Boyer J., Cattolico L., Confanioli F., de Daruvar A.,  
 RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,  
 RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,  
 RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,  
 RA Nicaud J.-M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,  
 RA Pellenz S., Potier S., Richard G.-P., Straub M.-L., Suleau A.,  
 RA Swennen D., Tekia F., Wesolowski-Louvel M., Westhof E., Wirth B.,  
 RA Zentou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,  
 RA Bouchier C., Caudron B., Scarpelli C., Gallardin C., Weissenbach J.,  
 RA Wincker P., Souciet J.-L.;  
 RT "Genome evolution in yeasts."  
 RL Nature 430:35-44 (2004).  
 DR EMBL; CR382131; CAG79360.1; -; Genomic DNA.  
 DR GO; GO:0008757; F:S-adenosylmethionine-dependent methyltransferase; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR InterPro; IPR000051; SAM\_bind.



Db 704 LRSASTKFWLTKYKNGLFQFEEENVRVQPGSSIDRVNMYAN 744

RESULT 6

Q6CJ18\_KJULA PRELIMINARY; PRT; 790 AA.

AC Q6CJ18; 2004 (TrEMBLrel. 28, Created)

DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)

DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)

DE Similar to ca|C1783|ipf7635 Candida albicans unknown function.

GN OrderedLocNames=KLLA0F22198;

OS Kluyveromyces lactis (Yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.

OX NCBI\_TaxID=28985;

RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].

RP STRAIN=CBS 2359 / IFO 1267 / NRRL Y-1140 / WM37;

RX PubMed=15229592; DOI=10.1038/nature02579;

RA Lafontaine I., de Montigny J., March C., Neuvéglise C., Talla E.,

RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,

RA Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykasten C.,

RA Boisrame A., Boyer J., Cattolico L., Confanioleri F., de Daruvar A.,

RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,

RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,

RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,

RA Nicaud J.-M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,

RA Pellenz S., Potier S., Richard G.-P., Straub M.-L., Suleau A.,

RA Swennen D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,

RA Zeniou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,

RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,

RA Wincker P., Souciet J.-L.,

RT "Genome evolution in yeasts."

RL Nature 430:35-44(2004).

DR EMBL; CR382126; CAG98779.1; -; Genomic DNA.

DR GO; GO:0008757; F:adenosylmethionine-dependent methyltransf. . .; IEA.

DR GO; GO:0016740; F:transferase activity; IEA.

DR InterPro; IPR000051; SAM\_bind.

KW Complete proteome; Methyltransferase; Transferase.

SQ SEQUENCE 790 AA; 90212 MW; 25BAE985E9C5B0D1 CRC64;

Query Match 25.7%; Score 890.5; DB 2; Length 790;

Best Local Similarity 33.2%; Pred. No. 1.1e-63;

Matches 233; Conservative 94; Mismatches 241; Indels 133; Gaps 23;

QY 51 LESFYGPQAAP-----ARLAERSNLIWVDLGGGTGGENVDMADYI 92

DB 93 LEQFYKSAQLYDRTRGVLLQGRSTSLKLSHLSEKGNVWIDVGGTGFNISQWALLT 152

QY 93 DL-AKFSIYVVDLCHSLSEVAKKAKGKNNQVVEADAPAPPEGTATLITFSYSL 151

DB 153 NLDTFDKLYLDLSPLSEVAKRKCKEKGKNNVEICGDACDFEIPESQAQITFSYSL 212

QY 152 TMIPPHNVIDQACSYL-SQDGLVGVADFYVSGKY-----DLPLRQMPWSRRFFW 200

DB 213 SMIPSPYALDHAVALDLAKNGIISCDFGVNESHMLVGRWTGLGVNHRHLPWLPRTW 272

QY 201 RSIFDIDNIDIGERRAYLEOKLERVWEQNTQ-----GSIYPVWVL----- 241

DB 273 RLWFEFDKFLDPAARLEYLRFGTIKSLNCYNKLGKIPYITWLGCKNDHEQHLQARFV 332

QY 242 ----RAPPYVWIGR-----LPSVGHALHEERV-----ERP- 268

DB 333 ELAPTSYPLAPITTSASSNAQPMWKAMIALENSKKGLPYSLFYQKEHWRVYDEVNPE 392

QY 269 --MFPPTFLYTOSWEPDPEMVEINPKDTVLITLTSGGCNALN--LLVQAGAGVVSVD 324

DB 393 YSQFNKSYIYATWEDPREDVNIINLTQPEDTILATISAGDNILHYATLPNPKRTHGVDL 452

QY 325 NPAQSALLKKAIAOQLFEFDVQWLFQGVGHPRIEELYKKLAPPLSQTSHNFWSKRLW 384

Db 453 NPCQGHLELKLAAIRSLSTFTQLWQMGEGKIDRFNNILNKLAPYLSSNAFYW----- 507

QY 385 YFOHGL-----YYQGMGKLCWVLOC---LAVVLGLGKTVKRLANAPTVEEORRLW 432

DB 508 -FENGTKTFDPNGAGLYDTGFTK--WALRLAKVVKFVANLTDEVNMLCKAKTIEEORSIW 564

QY 433 DSNMLIHFKVNGGPKPLVWLFVKFVSLVLF-NKAVLWFGGVGPKGVKALIKADGPIENYI 491

DB 565 DKKI-----KPV--LFNRVVGKILVGNPLFLMSALGVPRNQ---AKMGSSSTLOYI 610

QY 492 ARTMDGVAENSHVRKONYFYVNTCLTKFLRDNCPYTLREAFATLK-----SGVVDNLTVS 547

DB 611 IDTLDPVIDNSLSDNDNYFYTLTKGRYSRSCPYLKEGFGKLSRSRSPESPLDRVLH 670

QY 548 TNPF---MEELKARTYTKVILMDHVDWLDMPVANELAE---CLAKQVAPGGIWIWRSASL 601

DB 671 TDTLKDVCERLSKKTVSIAIIMDMDFD-POGTDVDEIQLWALNLSRGRVLLRSASK 729

QY 602 SPYAEIOLKAGFDVRCIRRATOG-YMDRVNMYSSFYMARR 641

DB 730 SPWYIKNFKFGFSCKAVSARYPGKICIDRVNMYASTWVCQK 770

RESULT 7

Q52GF6\_MAGGR PRELIMINARY; PRT; 832 AA.

AC Q52GF6;

DT 13-SEP-2005 (TrEMBLrel. 31, Created)

DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)

DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)

DE Hypothetical protein.

GN ORFNames=MG01330.4;

OS Magnaporthe grisea 70-15.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

OC Sordariomycetes incertae sedis; Magnaportheaceae; Magnaporthe.

OX NCBI\_TaxID=242507;

RN NUCLEOTIDE SEQUENCE.

RP STRAIN=70-15;

RA Birren B., Nusbaum C., Abebe A., Abouelleil A., Adekoya E.,

RA Ait-zahra M., Allen C., Allen T., An P., Anderson M., Anderson S.,

RA Arachchi H., Armbruster J., Bachantsang P., Baldwin J., Barry A.,

RA Bayul T., Blitshsteyn B., Bloom T., Blye J., Boguslavskiy L.,

RA Borowsky M., Boukhgalter B., Brunache A., Butler J., Caliste N.,

RA Calvo S., Camarata J., Campo K., Chang J., Cheshatsang Y., Cicroen M.,

RA Collymore A., Considine T., Cook A., Cooke P., Corum B., Cuomo C.,

RA David R., Dawoe T., Degray S., Dodge S., Dooley K., Dorje P.,

RA Dorjee K., Dorris L., Duffey N., Dupes A., Elkins T., Engels R.,

RA Erickson J., Farina A., Faro S., Ferreira P., Fischer H.,

RA Fitzgerald M., Foley K., Gage D., Galagan J., Gearin G., Gnerre S.,

RA Gnitke A., Goyette A., Graham J., Grandbois E., Gyaltsen K., Hafez N.,

RA Hagopian D., Hagos B., Hall J., Hatcher B., Heller A., Higgins H.,

RA Honan T., Horn A., Houde N., Hughes L., Hulme W., Husby E., Iliev I.,

RA Jaffe D., Jones C., Kamal M., Kamat A., Kamysseles M., Karlsson E.,

RA Kella C., Kieu A., Klesner P., Kodira C., Kulbokas E., Labutti K.,

RA Lama D., Landers T., Leger J., Levine S., Lewis D., Lewis T.,

RA Lindblad-toh K., Liu X., Lokyitsang T., Lokyitsang Y., Lucien O.,

RA Lui A., Ma L.J., Mabbitt R., Macdonald J., Maclean C., Major J.,

RA Manning J., Marabella R., Maru K., Matthews C., Mauceli E.,

RA McCarthy M., McDonough S., Mcghee T., Meldrim J., Meneus L.,

RA Mesrirov J., Mihalev A., Mihova T., Mikelsen T., Mlenga V., Moru K.,

RA Mozes J., Mulrain L., Munson G., Naylor J., Neues C., Nguyen C.,

RA Nguyen N., Nguyen T., Nicol R., Nielsen C., Nizzari M., Norbu C.,

RA Norbu N., O'donnell P., Okoawo O., O'leary S., Omotosho B.,

RA O'neill K., Oman S., Parker S., Perrin D., Phunkhang P., Pignani B.,

RA Purcell S., Rachupka T., Ramasamy U., Rameau R., Ray V., Raymond C.,

RA Retta R., Richardson S., Rise C., Rodriguez J., Rogers J., Rogov P.,

RA Rutnan M., Schupbach R., Seaman C., Settipalli S., Sharpe T.,

RA Sheridan J., Sherpa N., Shi J., Smirnov S., Smith C., Sougnez C.,

RA Spencer B., Stalker J., Stange-thomann N., Stavropoulos S.,

RA Steenson K., Stone C., Stone S., Stubbs M., Talamas J., Tchuinga P.,

RA Tensing P., Tesfaye S., Theodore J., Thoulutseang Y., Topham K.,

RA Towey S., Tsamla T., Taomo N., Vallee D., Vassiliev H.,

RA Venkataraman V., Vinson J., Vo A., Wade C., Wang S., Wangchuk T.,  
RA Wangdi T., Whittaker C., Wilkinson J., Wu Y., Wyman D., Yadav S.,  
RA Yang S., Yeager S., Yee E., Young G., Zainoun J., Zembeck L.,  
RA Zimmer A., Zody M., Lander E.,  
RT "The genome sequence of Magnaporthe grisea."  
RT Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=70-15;  
RA Dean R., Mitchell T., Brown D., Pan H., Thon M.;  
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=70-15;  
RA Zhu H., Blackmon B.;  
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.  
CC -!- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
DR EMBL; AACU01000030; EAA55679.1; -; Genomic\_DNA.  
KW Hypothetical protein; Methyltransferase; Transferase.  
SQ SEQUENCE 832 AA; 93320 MW; 98A5CF0B54A618AE CRC64;  
  
Query Match 25.5%; Score 881.5; DB 2; Length 832;  
Best Local Similarity 31.3%; Pred. No. 6.3e-63;  
Matches 226; Conservative 114; Mismatches 226; Indels 155; Gaps 21;  
  
QY 51 LESFYGPQAAAF-----AARLAERS-----NLIVDLGGGTGENDV 86  
DB 67 LESFYEQASIVDTTLLKGRDMLAALAAQLHRAKQSHQKRIWVDIGGTTGNIE 126  
QY 87 MMADYIDLAK-FKSIYVVDLCHSLCEVAKKAKGKKNVVEADACQF-----APP 138  
DB 127 AMRNFVPEFFSVVLVDPSLCEVARFORLGNKRVVVCQDARKFRIEDHEASGP 186  
QY 139 EG-----TATLIYFSYSLTMIPFPHNVIDQACSYLSQDGLGVADFYVSGKYD----- 186  
DB 187 KGNWQAIGADLLTWSYLSLMIPIFYSDLSLLAPGGIMGVVDFYVQSKVDPSFRNHT 246  
QY 187 --LPLRQMPWSRRFFWRSIFDINIDIGPERRAYLQKLERWEQNTQ-----GSIYYPVW 240  
DB 247 AGLTGRHCGFILRSFWRAFWDLDLSLEPARDYLFYRFTGLVGVNTRNRLGNI----- 301  
QY 241 LRAPYVWIG-----RLEPSV-----GHA---L 259  
DB 302 ---PYIWIICSTSGQSTQSLHENVSPKAVQVAVKNSAGLPLPSFYQNHAWRIY 358  
QY 260 HEBRVERPPMPFPFLYQTSWEDPEPDMYMEINPKDVTLTLSGGCNALNLLVQAGQV 319  
DB 359 YDQLRKHKQFNDEYIAFTWEDTRVDARLLKISSDDVLAITSAGDNLISYISQSPARV 418  
QY 320 VSDCHNPAQALLLEKVAIQLEFEDVQLFQEGVHPRIEELYKCLAPFLSQTSHFW 379  
DB 419 HAVDLNPSQNHLLLEKLASYSDALDYADFWKIFGEGHDEFRALLINKLSPLHSRAFYW 478  
QY 380 --SKRLWYFOGLYQYQGMKLCWVLOCLAVLGLGTVKRLANAPTMEQORLWDSNML 437  
DB 479 FANAHTSKHGXYDTGGSKHATRAFRWISSAFGVRSAVREMLEAKTINEQREIWOGR1- 537  
QY 438 IHFVKNPKPLVMLFVKFVS-LVLFNKAVLWFGGVGPKQYALIKADGI----- 485  
DB 538 -----RA--LLSRMVSNFVVSQQLWSALGVFKHQLAMLERDHATSKVVSERGEAS 587  
QY 486 -----PIENYIARTMDGVAENSHVRKONYFYNNCLTKOKFLRDNCPYVLRFAAFATL--KS 538  
DB 588 NSRTSAIQWMSDPLDPVAHDTHIAEENPYIYVVTMAGKFSRCHPDYLSPPAAHAKLSRRG 647  
QY 539 GVVDNLTVST---NFFMEELKARTYTKVILMDHVDWLDMP---VANELACLAKQVAPGG 592  
DB 648 ALERDRIHTEEDIVDLAGLAPSLTAVVWDSMDWFDPKPEPGVAAAQISKLNALPKGG 707  
QY 593 IVIWRSASLSPPAELIQKAGF-----DVCIRRRATQGYMDRVNMYSSFYMARRK 642

DB 708 RYLLRSAGLKPWYVEEFAKLGFPAPKRVGCRSPETRCI-----DRVNMVASCWLLTKT 759  
QY 643 G 643  
DB 760 G 760  
  
RESULT 8  
Q55SJ0\_CRYNE PRELIMINARY; PRT; 820 AA.  
AC Q55SJ0;  
DT 13-SEP-2005 (TREMELrel. 31, Created)  
DT 13-SEP-2005 (TREMELrel. 31, Last sequence update)  
DT 13-SEP-2005 (TREMELrel. 31, Last annotation update)  
DE Hypothetical protein.  
GN ORFNames=CNBE1930;  
OS Cryptococcus neoformans var. neoformans B-3501A.  
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Heterobasidiomycetes;  
OC Tremellomycetidae; Tremellales; Tremellaceae; Filobasidiella.  
OX NCBI\_TaxID=283643;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=B-3501A;  
RA Fung E., Hyman R.W., Rowley D., Bruno D., Miranda M., Fukushima M.,  
RA Wickes B.L., Fu J., Davis R.W.;  
RT "Cryptococcus neoformans serotype D sequencing."  
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.  
CC -!- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
DR EMBL; AAEY01000024; EAL20831.1; -; Genomic\_DNA.  
KW Hypothetical protein.  
SQ SEQUENCE 820 AA; 93328 MW; 0F694988B89353A2 CRC64;  
  
Query Match 25.4%; Score 881; DB 2; Length 820;  
Best Local Similarity 29.2%; Pred. No. 6.7e-63;  
Matches 223; Conservative 110; Mismatches 232; Indels 198; Gaps 21;  
  
QY 40 GSKGDDHAARLESFYGPQAAAF-----AARL-----AERS 70  
DB 88 GSRQKDH---LEAFVAGQADLYDTRSQLKGRETMLQLLAHLAKAQMVLPLVSAPSK 144  
QY 71 NLIVDLGGGTGENDVMADYIDLAKFKSIYVVDLCHSLCEVAKKAKGKKNVQVVEA 130  
DB 145 PRIWDLGGGTGWNIEKMDYLPITYFDIYIIDLCEPLEVARARIKARGKKNVHLCQ 204  
QY 131 DACQAPPGETATLIYFSYSLTMIPFPHNVIDQACSYL-SODGLGVADFYVVS---GKYD 186  
DB 205 DASRFVLPE-----WESGAIPPFYQVLDRCQVLDTPQGLMAVVDFTTGREVNKE 255  
QY 187 LPL-----RQMPWSRRFFWRSIFDINIDIGPERRAYLQKLERWEQNTQSGSIYYPVWLR 242  
DB 256 RAIGTASRKSWSFKWFWECWFDLDGVHLGSRRELYEKMGTIKYNAENNFINTWFIQ 315  
QY 243 APYVWIG----- 250  
DB 316 IPYVFLGCSQRDASASAKSFTEAGNRLGQSDGLLTPTSPFTNSPSVFGSPSPMLEM 375  
QY 251 -----RLPSVCHALHEERVERPPFPPTFLY 276  
DB 376 PELVLGFSAAQATQTFEVGAPLSPPHYHLRKWRIPYLEEKTHBQF-----RTHIY 427  
QY 277 TQSWEDPEPDMYMEINPKDVTLTLSGGCNALN-LLVQAGAGVSVVDCNPAQSALLLELK 335  
DB 428 GWTWEDPAVDVKLNKINDHIIAITSAGDNLVHYALTAKPARIHADVMNPPCQHILELK 487  
QY 336 KVAIQOLEFEDVWQLFQEGVHPRIEELYKCLAPFLSQTSHNFWSKLWTFQHLGYQGG 395  
DB 488 LAALQALEYNDFWLIFGEGRHPFEDKLLTTKLSPFLSSHAYAYWKTHTSQFSRNFYFRGY 547  
QY 396 MGKLCWVLOCLAV---VLGLGTVKRLANAPTMEQORLWDSNMLIHFVKNPKPLWLP 452



Db 548 SG---WALRLAQTAFFIAGVRGDKVRLCQATSTABQRIWQKI-----RP-VFLN 594  
 Qy 453 VKFVSLVFNKAVLWFGGGVPGKQYALIKADGPIENIARTMDGVAENSHVRKQNYFY 512  
 Db 595 KVMVGLFLGNPLFNWALGVQNMNCFLDG-SVEDYVKATLDPITLSTLKDNNVFL 653  
 Qy 513 NCLTGKFLRNCCTYLREAAAFATLK-SGVVDNLTVTSTNFMWELKA---RTYTKVILMDH 568  
 Db 654 LCLNGRYTRTSCPAFLKPEGFKALRNSKSTDAFKLHTDTTLNLRGLPDBESLTKIIVMS 713  
 Qy 569 VDWLD-----MPVAN-----ELAECLAKOVAAPGVIVIR 597  
 Db 714 MOWFEDIPCTPDPQDSTALDTLOATPEKALHRLAEIDYEIL-MKRVKVGGLIAVR 772  
 Qy 598 SASLSPPYAEILIOKAGPDRVCIR-RATQGYMDRVNMYSSFYMA 639  
 Db 773 SAAKRWYRQRFPAAGLKVPIDIRENQEAIDRVNMYVSPWKA 815

RESULT 9  
 ID Q6BK65 DEBHA PRELIMINARY; PRT; 763 AA.  
 AC Q6BK65;  
 DT 25-OCT-2004 (T-EMBLrel. 28, Created)  
 DT 25-OCT-2004 (T-EMBLrel. 28, Last sequence update)  
 DE Similar to CA1783|IPF7635 Candida albicans IPF7635.  
 GN OrderedLocNames=DEHA0F25982g;  
 OS Debaryomyces hansenii (Yeast) (Torulaspora hansenii).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Debaryomycetes.  
 OX NCBI\_TaxID=4959;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RC STRAIN=ATCC 36239 / CBS 767;  
 RX PubMed=15229592; DOI=10.1038/nature02579;  
 RA Dujon B., Sherman D., Fischer G., Durrans P., Casaregola S.,  
 RA Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,  
 RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,  
 RA Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykasten C.,  
 RA Boisrame A., Boyer J., Cattolico L., Confaniolero F., de Daruvar A.,  
 RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,  
 RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,  
 RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,  
 RA Nicaud J.-M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,  
 RA Pellenz S., Potier S., Richard G.-P., Straub M.-L., Suleau A.,  
 RA Swennen D., Tekia F., Wesolowski-Louvel M., Westhof E., Wirth B.,  
 RA Zenitou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,  
 RA Bouchier C., Caudron B., Scarpelli C., Gallardin C., Weissenbach J.,  
 RA "Genome evolution in yeasts."  
 RT Wincker P., Souciet J.-L.;  
 RL Nature 430:35-44(2004).  
 DR EMBL: CR382138; CG89817.1; -; Genomic DNA.  
 DR GO: GO:0008757; F1S-adenosylmethionine-dependent methyltransf. .; IEA.  
 DR GO: GO:0016740; F1S-transferase activity; IEA.  
 DR InterPro: IPR000223; Peptidase\_S26A.  
 DR InterPro: IPR000051; SAM\_bind.  
 DR PROSITE: PS00501; SPASE\_1; UNKNOWN 1.  
 KW Complete proteome; Methyltransferase; Transferase.  
 SQ SEQUENCE 763 AA; 87363 MW; 7979A2C300525F56 CRC64;

Query Match 25.0%; Score 864.5; DB 2; Length 763;  
 Best Local Similarity 31.4%; Pred. No. 1.4e-61;  
 Matches 223; Conservative 105; Mismatches 252; Indels 131; Gaps 20;

Qy 41 SKGDDHARLESFYGPQAAAF-----AARLARSNLIWVDLGGGTG 82  
 Db 57 TTKPTQQQSLELFYKQAHVYDKTRVLLKGRKECLRLATAHLKSKKDLVWVDIGGGTG 116  
 Qy 83 ENVMWADYIDLAK-FKSIYVVDLCHSLCEVAKKAKAGKGNVQVVEADACAFPEGT 141  
 Db 117 SNIEYMDQWMSISKFNKAVYLDVLSLCEVAKKRFADKRWNSVHVLVADACDFGIYBK 176

Qy 142 ATLITSYSILTMIPPHNVIDQACSYLSQDGLVGVAADFVYS-----GKYD-----LPLRQ 191  
 Db 177 ADLVTSYSLSIMPTFHGAVDHAKLLDKTGVIAICVDVGQTBESSVGRVITLGGIVNRN 236  
 Qy 192 MPMSRFFRSIFDIDNIDIGPERRAYLEOKLRVWEQNTQGSIPYVPMWRAPVYWG- 250  
 Db 237 IPVLRNFRWIFEADKVFLDLSRRHYLYKFKGTIKSLCYN- - - - - KLGRIPYIYVWGC 292  
 Qy 251 - - - - - RLP-SVGH- - - - - 257  
 Db 293 DKSSKSSLLNRINCLATESPYLAPSDKNKIDIPISKGHEAALMNFQKLPYPSIYQ 352  
 Qy 258 - - - - - ALHERVERPMPPTFLYTQSWEDPDPMEVWEINPKDTVLTLTSGGNALN--L 311  
 Db 353 EIWRVYFDEIDQYLOFQNYVYAFWEDPREPHNLKFTSEDTVLTAITSAGNLSYAT 412  
 Qy 312 LVQAGQVVSVDNCPAQSALLEKKAIVAOOLEPEDVWOLFGEVGHPRIELEYEKCLAPPL 371  
 Db 413 LNPFPKRIHADVLDNPCQNHLLLEKLKASFRALSKNQIWSIFGEGKIKNFKDLLVDKLSPHV 472  
 Qy 372 SQTSHNFW---SKRLWYFQHGLEYQGGMGKLCWLQC---LAVVLGLGKTVKRLANAPT 424  
 Db 473 SSNTFOYMDRGDKTFNPKGKGLY---DTGSTRWALRLAKWPKICGITKYVDMLCSDT 529  
 Qy 425 MBEORLWDSNMLIHFKVNGPKPLVWLFKVY-SLVLFNKAVLWFGGVPKQYALIKAD 483  
 Db 530 LBEQQRINWNI-----RPT--LFPNPFVSSLLIGNPIFLWKALGVPAQNAAM- 575  
 Qy 484 GPIENVIARTMDGVAENSHVRKQNYFYNCLTGKFLRNCCTYLREAAFATL- - - - - 536  
 Db 576 GNSILKYVVDTPVPIKRLISEDNIFYTLTKGKTPQPCPDYLTKEGKYSUTTNRKS 635  
 Qy 537 KSGVVDNLTVST---NFFMEELKARTYTKVILMDHVDLMDPMPVANELAECLAKQ--VAPG 591  
 Db 636 KEAPIDNIRLHTDMLNDVFARLSKLSLAIIMDHMDWDFDNGEDALNEITALKSLNTN 695  
 Qy 592 GIVVHRSASLSPYAEILIOKAGPDRVCIR-RATQGYMDRVNMYSSFYMAR 641  
 Db 696 GRVLLRASASTNPWYIKTFEDLQFTCKAAGIRDTGISIDRINMYASTVWCTK 746

RESULT 10  
 ID Q7SGY5 NEUCR PRELIMINARY; PRT; 908 AA.  
 AC Q7SGY5;  
 DT 01-MAR-2004 (T-EMBLrel. 26, Created)  
 DT 01-MAR-2004 (T-EMBLrel. 26, Last sequence update)  
 DT 01-FEB-2005 (T-EMBLrel. 29, Last annotation update)  
 DE Hypothetical protein (Related to S-adenosylmethionine:diacylglycerol  
 DE 3-amino-3-carboxypropyl transferase braa).  
 GN Name=NCU03032.1; Synonyms=B18P24.050;  
 OS Neurospora crassa.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.  
 OX NCBI\_TaxID=5141;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=OR74A;  
 RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,  
 RA Jaffe D., Fitzhugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,  
 RA Ekins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,  
 RA Qui D., Ianakiev P., Pedersen D., Nelson M., Washburne M.,  
 RA Selitreniukoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schulte U.,  
 RA DeSouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R.,  
 RA Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gnerre S.,  
 RA Kamal M., Kamysisselis M., Muceli E., Bielke C., Rudd S., Frisman D.,  
 RA Krystofova S., Ramussen C., Metznerberg R.L., Perkins D.D., Kroken S.,  
 RA Cogoni C., Macino G., Catchside D., Li W., Pratt R.J., Omani S.A.,  
 RA DeSouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R.,  
 RA Varden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,  
 RA Natvig D.O., Alex L.A., Mannhaupt G., Eboole D.J., Freitag M.,  
 RA Paulsen I., Sachs M.S., Lander E.S., Nusbaum C., Birren B.,  
 RT "The Genome Sequence of the Filamentous Fungus Neurospora crassa."

```
RL Nature 0:0-0(2003).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Schulte U., Aign V., Hoehsel J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.;
RL Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RA German Neurospora genome project;
RL Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABX01000006; EAA36182.1; -; Genomic DNA.
DR EMBL; BX842626; CAE76263.1; -; Genomic DNA.
DR GO; 0008757; F1S-adenosylmethionine-dependent methyltransf. .; IEA.
DR GO; 0016740; F1S-transferase activity; IEA.
DR InterPro; IPR000051; SAM bind.
KW Hypothetical protein; Methyltransferase.
SQ SEQUENCE 908 AA; 101376 MW; C20AC06831AA4154 CRC64;

Query Match 24.6%; Score 850.5; DB 2; Length 908;
Best Local Similarity 30.4%; Pred. No. 2.5e-60;
Matches 222; Conservative 105; Mismatches 235; Indels 169; Gaps 18;

QY 69 RSNLIWDLGGTGGNGVDMADYIDLAK-FKSYIVVDLCHSLCEVAKKAKAKGKNVOV 127
DB 178 RRPDIWVDDGGTGWNIEMAKFVNVSFFKTVILVDFSPSLCEVAKKPARLGWNRV 237
QY 128 VEADACQFA-----PPEGT-----ATLITFSYSLTWIPFP 157
DB 238 ICTDARKFLEDYEDVDEGSGSDSPSLSGMWGKTPGRHAGAEIITWSYLSMWPDY 297
QY 158 HNVIDQACSVLSDGLVGADVFVSGKYD-----LPLRMPWSRRFFWTSIFDIDN 208
DB 298 FSIIDSLESILAPHLIAVVDFAQSKVDPTFRNYTGGLMNRHVGFARNFWSWFDADR 357
QY 209 IDGIPERRAYLEOKLERVW---EONTGSIIPVPM---LRAPYVW-----247
DB 358 VSLPARDYLVFRGTVTVNARNVTLGAIPYVILGCLLKPFSTSSLPHEIVEHIDAI 417
QY 248 -----WIGR-----LPSVGHALH 260
DB 418 ATSPSSRPLVGHKSSATNALAFAGRTAPEMRSKAFNTAENISANLPLSFYQNH 477
QY 261 -----EERVERPMPPTLYTQSHEDPEPDMEVMEINPKDTVILTSCGCNALMLVQ 314
DB 478 HWRIYDDQLPKHTQFNDEYIYFTWEDSRVDRLLMLGPDVVLAITSGDNLISYLMQ 537
QY 315 GAGQVSVDCNPAQSALLEKVAIQOLEFEDVWOLFGEVGHPRIEBELYEKLPFLSQT 374
DB 538 SPARVHAIDNPAQNHLLELUKVASFTTLDTPDWKLFGEKGHPDFRSLLSKUSPLSGR 597
QY 375 SHNFWSKRLMYFO-----HGLYQGGMGKLCWLQCLAVVLGLGKTVKRLANAPTMBEQR 430
DB 598 AFQYWLNAHIFTDPAGRGLYDTGGSRYAIRFRFWISTLFFCRSAVERLLSTPLEQRS 657
QY 431 LWDNSMLIHFKVNGPKPLVWLFVKFVS-LVLFNKAVLWFGGVPKGQYALIKAD-----483
DB 658 IYHTKI-----RPC--LLNRFVNGVLSSDAFLWSALGVPKQVAMIEADYHRSI 706
QY 484 -----GPIENVARTMDGVAENSHVRKQNYFYNYCLTGKFLRNCPTYL 529
DB 707 SSSTPSKSKPSRAEAILHYTSTLDPLVLTSHLASDNFYLVCLVQVTRQCHPDYLS 766
QY 530 EAAFATLKS-GVVDNLFTVSTNPFMEEL---KARTYTKVILMDHVDLMDP-----575
DB 767 PAHSHLSAPGAFGLRIHTDIEQVLARFQPGTLTVAVWVMSMDWDFDPPSPPEEKGRG 826
QY 576 VANELAECLAKQVAPGIVTWRSNLSPPYAEILQKAGFVRCI-----RRATQGYMDRV 630
DB 827 KAREQVRLNRALKVGGKLLRSAGVEPWVYRVFVEEGFGARRVCGRESRGQDQECIDRV 886
```

---

```
QY 631 NMYSFYMARR 641
DB 887 NMVASCWILEK 897

RESULT 11
Q5KGX9 CRYNE PRELIMINARY; PRT; 797 AA.
AC Q5KGX9;
DT 10-MAY-2005 (Tremblrel. 30, Created)
DT 10-MAY-2005 (Tremblrel. 30, Last sequence update)
DT 10-MAY-2005 (Tremblrel. 30, Last annotation update)
DE Hypothetical protein.
OS ORFNames=CNE01970;
GN Cryptococcus neoformans var. neoformans JEC21.
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Heterobasidiomycetes;
OC Tremellomycetidae; Tremellales; Tremellaceae; Filobasidiella.
OX NCBI_TaxID=214684;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=JEC21;
RA Loftus B., Amedeo P., Roncaglia P., Vamathevan J., Utterback T.,
RA Van Aken S., Fraser C.;
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=JEC21;
RX PubMed=15653466; DOI=10.1126/science.1103773;
RA Loftus B.J., Fung E., Roncaglia P., Rowley D., Amedeo P., Bruno D.,
RA Vamathevan J., Miranda M., Anderson I.J., Fraser J.A., Allen J.E.,
RA Bosdet I.E., Brent M.R., Chiu R., Doering T.L., Donlin M.J.,
RA D'Souza C.A., Fox D.S., Grinberg V., Fu J., Fukushima M., Haas B.J.,
RA Huang J.C., Janbon G., Jones S.J.M., Koo H.L., Krzywinski M.I.,
RA Kwon-Chung K.J., Lengeler K.B., Maiti R., Marra M.A., Marra R.E.,
RA Mathewson C.A., Mitchell T.G., Pertea M., Riggs F.R., Salzberg S.L.,
RA Schein J.E., Svartsbeyn A., Shin H., Shumway M., Specht C.A.,
RA Suh B.B., Tenney A., Utterback T.R., Wickes B.L., Wortman J.R.,
RA Wye N.H., Kronstad J.W., Lodge J.K., Heitman J., Davis R.W.,
RA Fraser C.M., Hyman R.W.;
RT "The genome of the basidiomycetous yeast and human pathogen
RT Cryptococcus neoformans."
RL Science 307:1321-1324 (2005).
DR EMBL; AE017345; AAM43531.1; -; Genomic DNA.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 797 AA; 90388 MW; FD1082B7EB22B240 CRC64;

Query Match 24.4%; Score 845.5; DB 2; Length 797;
Best Local Similarity 29.1%; Pred. No. 5.3e-60;
Matches 221; Conservative 104; Mismatches 220; Indels 215; Gaps 21;

QY 40 GSKKGDHHAARLESFYGPQAAAF-----AARL-----AERS 70
DB 88 GSRDQKH---LEAFYAGQADLYDTRSQLKGRWMLQALLAAHLKAPWRLPVSAPSK 144
QY 71 NLIVDLGGGTGNVDMADYIDLAKFSIYVVDLCHSLCEVAKKAKAKGKNVQVEA 130
DB 145 PRWVLDGGGTGNVDMADYIDLAKFSIYVVDLCHSLCEVAKKAKGKNVHVLQ 204
QY 131 DACQAPPGETATLITFSYSLTWIPPHNVVIDAQSYL-SODGLVGADVFVS---GKYD 186
DB 205 DASRFVLPE-----WESGAIPFPYQVLDRCQDQVLDQGLMAVVDFTSREVGNKE 255
QY 187 LPL-----ROMPSRRFRFRSIFDIDNIDIGPERRAVLEOKLERVMEONTGSIYPVPLR 242
DB 256 RAIGTAKRVSWSFKWFEWCWFDLDGVHLHGSREYLEYKMGTKITYNARNFLNTWFIQ 315
QY 243 APYVWIG-----250
DB 316 IPYVVLGCSQRDASASAKSFTEAGNLGQSDGLGLLTPTSPFTNSPSVFGSPSPMLEM 375
QY 251 -----RLPSVCHALHEERVPFPFPPTFLY 276
DB 376 PELVLGFSAAQAQTITFEVGAPLSPPHYHLRKAWRIPYLEKIQF-----RTHIY 427
```



[1] NUCLEOTIDE SEQUENCE.  
 RC STRAIN=FGSC A4;  
 RA Birren B., Nusbaum C., Abouelleil A., Allen N., Anderson S.,  
 RA Arachchi H.M., Barna N., Bastien V., Bloom T., Boguslavskiy L.,  
 RA Boukgafter B., Butler J., Calvo S.E., Camarata J., Chang J.,  
 RA Choepel Y., Collymore A., Cook A., Cooke P., Corum B., Dearellano K.,  
 RA Dodge J.S., Dodge S., Dooley K., Dorris L., Elkins T., Engels R.,  
 RA Erickson J., Faro S., Ferreira P., Fitzgerald M., Gage D., Galagan J.,  
 RA Gardyna S., Gnerre S., Graham L., Grand-pierre N., Hafez N.,  
 RA Hagopian D., Hagos B., Hall J., Horton L., Hulme W., Iliev I.,  
 RA Jaffe D., Johnson R., Jones C., Kamal M., Kamat A., Karatas A.,  
 RA Kells C., Landers T., Levine R., Lindblad-Toh K., Liu G., Lui A.,  
 RA Ma L.-J., Mabbitt R., Maclean C., Macdonald P., Major J., Manning J.,  
 RA Matthews C., Mauceli E., McCarthy M., Meldrim J., Meneus L.,  
 RA Mihova T., Menga V., Murphy T., Naylor J., Nguyen C., Nicol R.,  
 RA Nielsen C.B., Norbu C., O'Connor T., O'Donnell P., O'Neill D.,  
 RA Oliver J., Peterson K., Phunkhang P., Pierre N., Parcell S.,  
 RA Rachupka A., Ramaaamy U., Raymond C., Retta R., Rise C., Rogov P.,  
 RA Roman J., Schauer S., Schupback R., Seaman S., Severy P., Smirnov S.,  
 RA Smith C., Spencer B., Stange-Thomann N., Stojanovic N., Stubbs M.,  
 RA Talamas J., Tencave S., Theodore J., Topham K., Travers M.,  
 RA Vassiliev H., Venkataraman V.S., Viel R., Vo A., Wang S., Wilson B.,  
 RA Wu X., Wyman D., Young G., Zairoun J., Zemlek L., Zimmer A., Zody M.,  
 RA Lander E.;  
 RA "Genome Sequence of Aspergillus nidulans.";  
 RT Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AACD01000101; EAA57777.1; -; Genomic DNA.  
 DR GO: GO:0004719; Fibrinogen-D-aspartate (D-aspartate) O-meth. .; IEA.  
 DR GO: GO:0016740; Fibrinogenase activity; IEA.  
 DR GO: GO:0006464; Protein modification; IEA.  
 DR InterPro: IPR000682; PCMT.  
 DR InterPro: IPR000051; SAM\_bind.  
 KW Hypothetical protein.  
 SQ SEQUENCE 790 AA; 90079 MW; C3D808C2EC30C7B CRC64;  
  
 Query Match 21.7%; Score 752; DB 2; Length 790;  
 Best Local Similarity 29.0%; Pred. No. 2.4e-52;  
 Matches 200; Conservative 101; Mismatches 236; Indels 153; Gaps 19;  
  
 QY 69 RSNLIWDLGGGTGE--NVDMDYIDLAKFYS--TYVDLCHSLCEVAKKAKAGKNV 125  
 DB 116 KGXAIWDLGGGTGE--NVDMDYIDLAKFYS--TYVDLCHSLCEVAKKAKAGKNV 175  
  
 QY 126 QVVEADACOPAPPE-----GTATITFSYSLTMTPPHNVIDQACSVLSQDGLV 175  
 DB 176 TVVQDARAFQLEPDHIDPLKSVGAGADLVTVSYSLMT--PVQSIDVVS----- 223  
  
 QY 176 VADFYVSGKYDLPLRQMPRSRRFFWRSIFDINIDIGPERRAYLQKLERVWQNTQ--- 232  
 DB 224 -SRNYIGGVFN---RHVNWLGRAFWEAFWEADRVNLDAAARDYLEVRFCTVISASERNYL 279  
  
 QY 233 -GSIYPVPLRABYYWIGR----- 251  
 DB 280 LGGI-----PYTIFIGRQKDIYPNQASREAIEKLDASFTESPYLSPANHREEMNNAI 331  
  
 QY 252 -----LPSVGHALHEERV-----ERPPMPPTFLTQSWE 281  
 DB 332 IQSTQEIHSKAYSASVNVLSANLPLPAAYQNHHRIFVNDLLPKHTQFKNEYIYAFNWE 391  
  
 QY 282 DPDPDMEVMEINPKDVTLTITSGGCNALLVQAGQVQVSDNCNPAQSALELKKVATQ 341  
 DB 392 DPRVDHRLNLIKRDYVLAITSAGNILDYQKSPRRVHVDLNFQNHILKVASPIA 451  
  
 QY 342 LEFEDVWQLFGEGVHPRIBELYKKLAPFLQTSHTNFWSKRLWYFQ-----HGLYYQGMG 397  
 DB 452 LGHRDVKWIFEGCKHPFERNLITSRSLPSLSSQAFQYWLHSHVFTSSSGRLYETGSR 511  
  
 QY 398 KLCVWLQCLAVVLGLGTVKVKRLANAPTEBEEQRLDSDNMLIHFKVNGPKPLVFLVKFVS 457  
 DB 512 HATKMIKRIYFKVFLGEGQVKUCEQTALBQKQIWPTRAVILL-----SKPLHW----- 560  
  
 QY 458 LVLFNKAFLVFGGVFGKQYALIKAD-----GIPIENTIARTMDGVA 499

2000

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2006, 01:53:31 ; Search time 104 Seconds  
(without alignments)  
2737.668 Million cell updates/sec

Title: US-10-620-914-45  
Perfect score: 3463  
Sequence: 1 MSGGRDPASVTKNFKSLE.....RVNYSSTFMMARKKAKXN 648

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_21.\*  
1: Geneseq1980s:\*  
2: Geneseq1990s:\*  
3: Geneseq2000s:\*  
4: Geneseq2001s:\*  
5: Geneseq2002s:\*  
6: Geneseq2003s:\*  
7: Geneseq2003bs:\*  
8: Geneseq2004s:\*  
9: Geneseq2005s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3463	100.0	648	9	Adw72748 Chlamydom
2	891	25.7	752	8	Adp98846 C. albica
3	850.5	24.6	908	9	Adw72753 Neurospor
4	842.5	24.3	845	8	Adr86056 Aspergill
5	381.5	11.0	416	9	Adw72756 Rhodobact
6	379.5	11.0	416	6	Abg72176 Rhodobact
7	379.5	11.0	416	9	Adw72706 Rhodobact
8	326.5	9.4	416	6	Abg72178 Agrobacte
9	326.5	9.4	416	9	Adw72732 Agrobacte
10	321	9.3	415	9	Adw72744 Mesorhizo
11	292.5	8.4	416	6	Abg72180 Sinorhizo
12	292.5	8.4	416	9	Adw72736 Sinorhizo
13	123	3.6	2284	8	Adt04703 House mou
14	121	3.5	212	7	Adc97627 E. faeciu
15	120	3.5	221	6	Abg72181 Sinorhizo
16	120	3.5	221	9	Adw72738 Sinorhizo
17	119	3.4	2286	8	Adt04639 Human DNA
18	118	3.4	329	9	Abm97161 M. xanthu
19	112.5	3.2	672	6	ABU40254 Protein e
20	111	3.2	195	8	Adk48246 Streptoco
21	111	3.2	198	8	Adr95987 Novel S.
22	111	3.2	198	9	Aea59857 Streptoco
23	110.5	3.2	298	4	Abg07975 Novel hum
24	108.5	3.1	706	7	Ab076020 Pseudomon

25	107	3.1	255	4	AAB67844	Aab67844 Polypepti
26	107	3.1	255	4	AAm38861	Aam38861 Human pol
27	107	3.1	255	4	AAG67149	Aag67149 Amino aci
28	107	3.1	255	8	ABM80519	ABm80519 Tumour-as
29	107	3.1	259	4	AAm40647	Aam40647 Human pol
30	107	3.1	485	4	AAm41069	Aam41069 Human pol
31	106.5	3.1	670	6	ABU41847	ABu41847 Protein e
32	106	3.1	472	5	ABB78345	ABb78345 Amino aci
33	106	3.1	472	6	ABU03555	ABu03555 Angiogene
34	106	3.1	472	6	ABU57638	ABu57638 Different
35	106	3.1	472	7	ADR25782	Adr25782 Human pro
36	106	3.1	472	7	ADF76582	Adf76582 Novel hum
37	106	3.1	472	8	ABM81165	ABm81165 Tumour-as
38	106	3.1	2039	8	ADS43564	ADs43564 Bacterial
39	105.5	3.0	472	4	AAm39283	Aam39283 Human pol
40	105	3.0	234	2	AAr92059	Aar92059 Heptapren
41	104.5	3.0	663	6	ABU31973	ABu31973 Protein e
42	104.5	3.0	664	8	ADs14914	ADs14914 Pseudomon
43	104.5	3.0	677	7	ABO66571	ABo66571 Klebsiell
44	104	3.0	225	9	ADW72745	Adw72745 Mesorhizo
45	103.5	3.0	1070	7	ADC94089	Adc94089 E. faeciu

ALIGNMENTS

RESULT 1  
ADW72748  
ID ADW72748 standard; protein; 648 AA.  
XX  
AC ADW72748;  
XX  
DT 21-APR-2005 (first entry)  
XX  
DE Chlamydomonas reinhardtii Btal gene, protein.  
XX  
KW Phospholipid synthesis; transgenic plant; fertilizer; enzyme;  
KW betaine lipid.  
XX  
OS Chlamydomonas reinhardtii.  
XX  
PN WO2005009115-A2.  
XX  
PD 03-FEB-2005.  
XX  
PF 15-JUL-2004; 2004WO-US022789.  
XX  
PR 16-JUL-2003; 2003US-00620914.  
XX  
PA (UNMS ) UNIV MICHIGAN STATE.  
XX  
PI Benning C, Riekhof W;  
XX  
DR WPI; 2005-112975/12.  
DR N-PSDB; ADW72746, ADW72747.  
XX  
PT New composition comprising a purified DNA having an oligonucleotide  
sequence encoding a protein, useful in producing Betaine lipid compounds  
e.g., Diacylglycerol-O-4'-(N,N,N-trimethyl) homoserine (DGTS).  
XX  
PS Claim 11; SEQ ID NO 45; 147pp; English.  
XX  
CC The invention relates to a new composition comprising a purified DNA  
having an oligonucleotide sequence appearing as ADW72747 (Chlamydomonas  
reinhardtii Btal cDNA) or ADW72752 (Neurospora crassa Btal coding region,  
encoding Btal proteins which are DGTS (diacylglycerol-O-4'-(N,N,N,-  
trimethyl)homoserine) synthetic enzymes. Btal (and BtaA and BtaB) enzymes  
allow the replacement of phospholipids with non-phosphorus containing  
lipids in the cell membrane. A transgenic plant expressing the enzymes  
CC would have a reduced need for phosphate-containing fertilizer. Also  
included are an RNA transcribed from the purified DNA, antibodies  
produced from the protein, a vector comprising the DNA or nucleic acid, a  
host cell comprising the vector, transgenic plants comprising the vector,

CC and a protein translated from the RNA or encoded by the nucleic acid. The  
CC composition is useful in producing betaine lipid compounds e.g.,  
CC Diacylglycerol-0-4'-(N,N,N-trimethyl) homoserine (DGTS). The composition  
CC of the invention reduces the amount of phosphate fertilizer needed for  
CC the optimal growth of crop plants. The present sequence represents a  
CC betaine lipid synthetic enzymes.

Sequence 648 AA;

Query Match	100.0%	Score 3463;	DB 9;	Length 648;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 648;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;

Qy	1	MGSRDRPASYTKNFSLKXKLKLSMKDDTLVLRHMMFGSKGDDHAARLESFVGPQAA	60
Db	1	MGSRDRPASYTKNFSLKXKLKLSMKDDTLVLRHMMFGSKGDDHAARLESFVGPQAA	60
Qy	61	AFAARLAERSNLIWVDLGGGTGENVDMADYIDLAKFKSIYVVDLCHSLCEVAKKAKAK	120
Db	61	AFAARLAERSNLIWVDLGGGTGENVDMADYIDLAKFKSIYVVDLCHSLCEVAKKAKAK	120
Qy	121	GWKNVQVEADACQAFAPPEGATLITFSYSLTMTIPPHNVTDQACSYLSQDGLVGVADFY	180
Db	121	GWKNVQVEADACQAFAPPEGATLITFSYSLTMTIPPHNVTDQACSYLSQDGLVGVADFY	180
Qy	181	VSGKYDLPLOMPWSRRFFWESIYFDINIDIGPERRAYLEOKLBRVWEQNTQGSIPYVPW	240
Db	181	VSGKYDLPLOMPWSRRFFWESIYFDINIDIGPERRAYLEOKLBRVWEQNTQGSIPYVPW	240
Qy	241	LRAPYYVWIGRLPSVGHALHEERVPRPMPPTFLYTQSWBDDPDMVMEINPKDVTLT	300
Db	241	LRAPYYVWIGRLPSVGHALHEERVPRPMPPTFLYTQSWBDDPDMVMEINPKDVTLT	300
Qy	301	LTSGGCNALNLLVQAGQVSVDCNPAQSALLELKKVAIQOILEPDMWQLPBGVHPRIE	360
Db	301	LTSGGCNALNLLVQAGQVSVDCNPAQSALLELKKVAIQOILEPDMWQLPBGVHPRIE	360
Qy	361	ELYEKKLAPFLSQTSHNFWSKRLWYFQHGLYYQGGMGKLCWVLOCLAIVLGLGKTVKRLA	420
Db	361	ELYEKKLAPFLSQTSHNFWSKRLWYFQHGLYYQGGMGKLCWVLOCLAIVLGLGKTVKRLA	420
Qy	421	NAPTMESQRRLWDSNMLIHFYVKGKPKLIVLIFVKFVSLVLFNKAVLWFGGGVPGHQYALI	480
Db	421	NAPTMESQRRLWDSNMLIHFYVKGKPKLIVLIFVKFVSLVLFNKAVLWFGGGVPGHQYALI	480
Qy	481	KADGIPLENYIARTMDGVAENSHVRKQNYFYNYNCITGKFLRDNCTPYLREAAFAFLTKSGV	540
Db	481	KADGIPLENYIARTMDGVAENSHVRKQNYFYNYNCITGKFLRDNCTPYLREAAFAFLTKSGV	540
Qy	541	VDNLTVSTNFFMEELKARTYTKVILMDHVDLMDPVANELAECLAQVAPGGIWIWRAS	600
Db	541	VDNLTVSTNFFMEELKARTYTKVILMDHVDLMDPVANELAECLAQVAPGGIWIWRAS	600
Qy	601	LSPPYAELIQKAGDVRICIRATQGYMDRVNMYSSFYMARRKGAKKDN	648
Db	601	LSPPYAELIQKAGDVRICIRATQGYMDRVNMYSSFYMARRKGAKKDN	648

RESULT 2  
ADP98846

ID ADP98846 standard; protein; 752 AA.

ADP98846:

23-SEP-2004 (first entry)

DE C. albicans specific gene, orf6.3438, protein sequence.

XX Diploid fungal cell; allele; gene disruption cassette;  
KW promoter replacement fragment; antifungal; fungicide; gene therapy;  
XW infection; *Candida albicans*.  
KW  
XX  
OS *Candida albicans*.

or for preventing or inhibiting formation on a surface of a biofilm comprising *Candida albicans*; a pharmaceutical composition comprising a therapeutic amount of an agent which reduces the activity or level of a gene product encoded by a nucleic acid comprising any of ADP98516-ADP98825 in a pharmaceutical carrier; an antibody preparation which binds the polypeptide; methods for evaluating a compound against a target gene product encoded by any of ADP98516-ADP98825; identifying an antimycotic compound; a computer or a computer readable medium that comprises at least one of the nucleotide sequences mentioned in the specification or at least one amino acid sequence selected from ADP98826-ADP99135; a method assisted by a computer for identifying a putatively essential gene of a fungus; and a protein array comprising proteins, where at least one protein comprises an amino acid sequence or a portion of an amino acid sequence selected from ADP98516-ADP98825. The novel methods and compositions have fungicide activity. The compositions may be used in gene therapy. The composition and methods are useful for drug screening purposes or for diagnosing, preventing or treating infections associated with *Candida albicans*. These may also be used for constructing strains useful for identification and validation of gene products as effective targets for therapeutic intervention, for identifying and validating gene products as effective targets for therapeutic intervention, and for collecting identified essential genes. This sequence represents the protein of a *Candida albicans* fungal specific gene of the invention. CC NOTE: This sequence was downloaded from an electronic sequence listing provided on the WIPO website.

XX  
SQ Sequence 752 AA;

Query Match 25.7%; Score 891; DB 8; Length 752;  
Best Local Similarity 32.7%; Pred. No. 1.2e-82;  
Matches 229; Conservative 98; Mismatches 232; Indels 142; Gaps 21;

QY 51 LESFYGFQAAAF-----AARLAERSNLIVDLGGGTGENVDMADVI 92  
DB 70 LESFYKNQAHYDNTREFLKGRQECRLRAISHLPKKKLLIWDICGGTGSNIEFDEIS 129  
QY 93 DLAK-FKSIYVDLCHSLCEVAKKAKAKGKNVQVVEADACQAPPEGTATLITFSYL 151  
DB 130 KISENFKAVYLDVLSPLCEVAKARFEHWTNVHVLVADACDFTIDYSDADLITFSYL 189  
QY 152 TMIPPHNVIDQACYSLOGLVGVADFVYS-----GKYD-----LPLQMPWSRRFFWR 201  
DB 190 SMIPFTNAIDNAVSKLDEGIATVDFGQSSTSMGRINTVGGVLRNDRIDPILNFWR 249  
QY 202 SIFDIDNIDIGPERRAYLEOKLERVWEQNTQ-----GSIPVPLWLRAPYVYVIGRLPSVGH 257  
DB 250 IYFEADKVLDSRRRYLEYKFTVKSLSYNKALGI-----PYIWIQCDKSKSH 301  
QY 258 ALHEE-----RVERPPMPPT-----273  
DB 302 TILERLNLCLATESPYLAPTTPITIANQLEDIPISKGHEAALINLQKLPVPSMYQKEYWR 361  
QY 274 -----FLYTSQWEDPEPDMVMEINPKDTVLTLTSGGCNALML---LVQG 315  
DB 362 VYIDEMNPLYEQPKQYIYAFWEDPREDKLNFSTDDTTLAITSAGDNILSYASLPTP 421  
QY 316 AGOVSVDCNPAQSALLEKKAIVQLEFEDVWQLGEGVHPRIELEYKELAPLSQTS 375  
DB 422 PKKIHAVDLNCPQHLLLEKLASFRCLSQEQIWSMFGEGKIENFNOLLITDTPHNSNA 481  
QY 376 HNFWSKR--LWYFQHLGYTQGGMGKLCWLQC---LAVVLGLGKTVKRLANAPTMEEQRR 430  
DB 482 FQYWMKDKPFTSGKGLY---DTGFSRMARLSRYVFKVGVSKSYVEELCAATMEEQLR 538  
QY 431 LWDNMLIHFKVNGPKPLVWLFVKFV-SLVLFNKAVLWFGGVGVPKQYALIKADGPIEN 489  
DB 539 INNEHL-----KPT--LFPVVGSLVGVNPFMLWKLALGVPAQAALM---GPSVIK 584  
QY 490 YIARTWDGVAENSHVRKQNYFYNCILTKFLRNCPTYLREAFATLK-----SGVV 541  
DB 585 YVVDITLDPILKRSISMISNDNYFYFLCMGGRYTKNCPDYLTITTKFNRLSSTAATASGSSI 644  
QY 542 DNLTVSTNFFME---ELKARTYTKVILMDHVDWLD---MPVANELAECLAKQVAPGGIVI 595

DB 645 DNLRIHTDTLNEVFGRLKEKSIATIIIMDMWFDPNGRDAINEIT-ALKRCLAPGGRVL 703  
QY 596 WRSASLSPPYAEELQKAGF-DVRCIRRAATQGYMDRVNMYSS 635  
DB 704 LRSASTKPYWLKTFKMLGFGQEEENVVRQPGSSIDRVNMYAN 744

RESULT 3  
ADW72753  
ID ADW72753 standard; protein; 908 AA.  
XX  
AC ADW72753;  
XX  
DT 21-APR-2005 (first entry)  
XX  
DE *Neurospora crassa* Btal gene, protein.  
XX  
KW Phospholipid synthesis; transgenic plant; fertilizer; enzyme; betaine lipid.  
XX  
OS *Neurospora crassa*.  
XX  
FN WO2005009115-A2.  
XX  
PD 03-FEB-2005.  
XX  
PF 15-JUL-2004; 2004WO-US022789.  
XX  
PR 16-JUL-2003; 2003US-00620914.  
XX  
PA (UNMS ) UNIV MICHIGAN STATE.  
XX  
PI Benning C, Riekhof W;  
XX  
DR WPI; 2005-112975/12.  
XX  
PT N-PSDB; ADW72751, ADW72752.  
XX  
PT New composition comprising a purified DNA having an oligonucleotide sequence encoding a protein, useful in producing Betaine lipid compounds e.g., Diacylglycerol-O-4'-(N,N,N-trimethyl) homoserine (DGTS).  
XX  
PS Claim 11; SEQ ID NO 50; 147pp; English.  
XX  
CC The invention relates to a new composition comprising a purified DNA having an oligonucleotide sequence appearing as ADW72747 (*Chlamydomonas reinhardtii* Btal cDNA) or ADW72752 (*Neurospora crassa* Btal coding region, encoding Btal proteins which are DGTS (diacylglycerol-O-4'-(N,N,N-trimethyl)homoserine) synthetic enzymes. Btal (and BtaA and BtaB) enzymes allow the replacement of phospholipids with non-phosphorus containing lipids in the cell membrane. A transgenic plant expressing the enzymes would have a reduced need for phosphate-containing fertilizer. Also included are an RNA transcribed from the purified DNA, antibodies produced from the protein, a vector comprising the DNA or nucleic acid, a host cell comprising the vector, transgenic plants comprising the vector, and a protein translated from the RNA or encoded by the nucleic acid. The composition is useful in producing Betaine lipid compounds e.g., Diacylglycerol-O-4'-(N,N,N-trimethyl) homoserine (DGTS). The composition of the invention reduces the amount of phosphate fertilizer needed for the optimal growth of crop plants. The present sequence represents a betaine lipid synthetic enzyme.

XX  
SQ Sequence 908 AA;

Query Match 24.6%; Score 850.5; DB 9; Length 908;  
Best Local Similarity 30.4%; Pred. No. 2.8e-78;  
Matches 222; Conservative 105; Mismatches 235; Indels 169; Gaps 18;

QY 69 RSNLIVDLGGGTGENVDMADYIDLAK-FKSIYVVDLCHSLCEVAKKAKAGKNVQV 127  
DB 178 RKKPIVDVGGGTGMNIEAKFPVNVSEFKTVILVDFSLCEVCEVAPARLGNVRV 237  
QY 128 VEADACQFA-----PPEGT-----ATLITFSYSLTIPPF 157

Db 238 ICTDARKFRLEDYEDYDEGSGSDSPSLSGWGETKPGRHAGAEILITMSYLSWMPDY 297  
QY 158 HNVIDQACSVLSQGLGVADVFVSGKYD-----LPLROMPSRRFFWRSIFDIDN 208  
Db 298 FSIIDSLSLAPHLIAVDFVFAQSKVDFTFRNYTGGLMNRHVGFARNFWSWFDADR 357  
QY 209 IDIGPERRAYLEQKLERVW---EONTQSGIPYVPW---LRAPYVW----- 247  
Db 358 VSLPARRDLYRFGVTLTVARNNTLGAIPYIWLGLCKLPKFPSTSLPHEIVEHDAI 417  
QY 248 -----WIGR-----LPSVGHALH 260  
Db 418 ATEPSRSPRLVGKSSSATNALAFVGRTPAPENRKAFTAIENISANLPLSPFFYQNH 477  
QY 261 -----EERVERPMPPTPLTQSWEDPEPDMVEINPKDVTLTITSGCNALNLVQ 314  
Db 478 HWRYYDDQLPKHTQNDENYIYFTWEDSRVDRELLNLGDDVVLTAISAGDNLILMQ 537  
QY 315 GAGQVSVDCNPAQSALLELKKVAIQOLEPEDVWQLFGEGVHPRIEELYEKKLAPLSQT 374  
Db 538 SPARVHAIDLNPQNHLLEKVASFTLLDYPDWKIFGEGKHPDFRSLILSKLSPLHSGR 597  
QY 375 SHNFWKRLWFO----HGLYYQGGMGKLCWVLOCLAVLGLGKTVKRLANAPTMBEQR 430  
Db 598 AFQYWLNAHIFTDPAGRGLYDTGGRYAIRFRFWISTLFFCRSAVRRLISTPLEGQRS 657  
QY 431 LWDNSMLIHFKVNGPRLWLVFKVYS-LVLFNKAVLWFGGVGPKQYALIKAD----- 483  
Db 658 IYHTKI-----RPC--LLARFVNLVLSDDAFLLWSALGVFKQKQVAMIEADYHRSI 706  
QY 484 -----GPIENYIARTWDGVAENSHVRKQNYFYNCILTKGFLRNCPTYL 529  
Db 707 SSTTSSKEKPSRAEAILHYTTSTLDPVLSHLSADNPYILVCLVLTQYTRQCHPYLS 766  
QY 530 EAPATLKS-GVVDNLTVSNTFMEEL---KARTYTKVILMDHVDMLDMP----- 575  
Db 767 PAAHSILSAPGAFDGLRIHTDEIQEVLARPQPGTLTVAVVMDSDWDFDPPSPBEEKEGR 826  
QY 576 VANELAECLAKOVAPGIVTWRSASLSPPYAELIQAGFDVRCI-----RRATQGYMDRV 630  
Db 827 KAREQVRLNRALKVGGKVLRSAGVEFWYRVFVEEGFARRVGCRSGRGDQECIDRV 886  
QY 631 NMYSSFYMAR 641  
Db 887 NMYASWILEK 897

## RESULT 4

ID ADR86056  
XX ADR86056 standard; protein; 845 AA.  
AC ADR86056;  
XX  
DT 04-NOV-2004 (first entry)  
XX  
DE Aspergillus fumigatus essential gene protein #106.  
XX  
KW Fungicide; Aspergillus fumigatus infection; Farmer's lung disease;  
KW drug screening.

Aspergillus fumigatus.

W02004067709-A2.

12-AUG-2004.

16-JAN-2004; 2004WO-US001099.

17-JAN-2003; 2003US-0441281P.

13-JUN-2003; 2003US-0478196P.

(ELIT-) ELITRA PHARM INC.

PA (ELIT-) ELITRA CANADA LTD.  
PI Jiang B, Hu W, Lemieux S, Roemer T;  
XX WPI: 2004-594200/57.  
DR N-PSDB; ADR85469.  
XX New purified or isolated Aspergillus fumigatus nucleic acid molecule  
PT encoding a gene product, useful for diagnosing and/or treating invasive  
PT fungal infections, such as Farmer's lung disease.  
XX  
PS Claim 1; SEQ ID NO 3106; 164pp; English.  
XX  
CC The present invention relates to Aspergillus fumigatus genes that are  
CC essential and are potential targets for drug screening. The methods and  
CC compositions of the present invention are useful for diagnosing and/or  
CC treating invasive Aspergillus fumigatus infection, including the allergic  
CC forms of the disease, such as Farmer's lung disease. They can also be  
CC used in various drug discovery purposes, such as expression of the  
CC recombinant protein, hybridization assay and construction of nucleic  
CC arrays. The present sequence represents an Aspergillus fumigatus  
CC essential gene protein sequence, used during diagnosis and drug  
CC development in the invention. These genes share a high degree of sequence  
CC conservation with known essential genes of candida albicans. The sequence  
CC data for this patent is not represented in the printed specification, but  
CC was obtained in electronic format from WIPO.  
XX  
SQ Sequence 845 AA;

Query Match 24.3%; Score 842.5; DB 8; Length 845;  
Best Local Similarity 30.1%; Pred. No. 1.7e-77;  
Matches 230; Conservative 96; Mismatches 254; Indels 183; Gaps 21;  
QY 42 KKG-DDHAARLESFYGPQAAAF-----AARL-----AERSNLIW 74  
Db 69 KGLNQDQALSFYKQGVDTAKRLLCGRDMLGLVAAQLKYKVENKELOAGKAIW 128  
QY 75 VDLGGGTGENDMADYIDLAKFS-IYVVDLCHSLCEVAKKAKAKGKQVQVREADAC 133  
Db 129 VDGGGTGYNIEAMASFLPTQFHSVILVDLSPSLCEVARQERLGGKWNVTWVCQAR 188  
QY 134 QF-----APPEGTATLITFSYSLTMTIPPHNVIDQACSVLSQDGLGVADFYV 181  
Db 189 SFLPHEKVDPRAPAPSTAGADLITMSYLSLMDPYYSVVDLTLDRASGLGVCDFFV 248  
QY 182 SGKYDLPL-----ROMPSRRFFRSIFDIDNIDIGPERRAYLEQKLERVWEONTQ 232  
Db 249 QSIVDVSARNYTGAFNRHVNLGRAFWRAFADRVSLAEARDVLEYRFGVVISASER 308  
QY 233 ---GSIPYVPLRAPYVWIGR-----LPSVGHALHERV-----ERPMPFPFLTYQ 278  
Db 309 NYLLGGI-----PYIFVGRHKNLASNLGQETIEMLDASFTSPYLSPANHTKEME 360  
QY 252 -----LPSVGHALHERV-----ERPMPFPFLTYQ 278  
Db 361 KAMERNAQEIIRSKAYESAVINLSANLPLSPSSFYQNHHCRIFYNDLLPKFKNEYIYAF 420  
QY 279 SWEDPEPDMVEINPKDVTLTITSGCNALNLVQAGOVSVDCNPAQSALLELKKVA 338  
Db 421 NWEDPRVDHRLDIDKDDVILTAISAGDNLIDYLOKSPRVHADVLDNPNQNHLELKVAS 480  
QY 339 IOOLEPEDVWQLFGEGVHPRIEELYEKKLAPLSOTSHNFWSKRLWYFO----HGLYYQ 394  
Db 481 FMAIGHRDVWKIFGEGKHPFELLISRLSAHLSLSSQAFQWLEHTHTFTSKYKGKLYETG 540  
QY 395 GMKGLWLOCLAVLGLGKTVKRLANAPTMBEQRRLWDSNMLIHFKVNGPKPLVWLTVK 454  
Db 541 GSRHAIKMVRYLEFKVFGLEGQVKLCCEAQTLABQREIWPKIRAVLM-----SKPLHW----- 592  
QY 455 FVSLVLFNKAVLWFGGVGPKQYALIKAD-----GPIENYIARTMD 496  
Db 593 ---AVSTEFWFAKAGVPRNQRNMLVDDYFKLGLTKDNQKGDISGRSIWQVVDLTD 649

QY 497 GVAENSHVRKQNYFYNYCLTGKELR-----DNC-----PTYLREAPATILKS-GVVDNLT 545  
 Db 650 PVNVTNISMNDNYFYCLCQSQFSRRVQTNQSKCIRDPTLYSPQARHVKLSSPGAFDGLR 709  
 QY 546 VSTNPFMEELK---ARTYTKVILMDHVDLMDMPVANEL---AECLAKQVAPGGIVWRSA 599  
 Db 710 IHTDEINEVKRTPSLTITAVIMDSXWFD-PEGTEASTQAKLNHALKMDGRILLRSA 768  
 QY 600 SLSPPYABLIQKAGFVRCI-RRATOGYMDRVNMYSSFYMARR 641  
 Db 769 SIEPFWYKQPEENGFTARRVGARFPGSGCIDRVNMYASTWICTK 811

## RESULT 5

ADW72756  
 ID ADW72756 standard; protein; 416 AA.

XX AC ADW72756;  
 XX DT 21-APR-2005 (first entry)  
 XX Rhodobacter sphaeroideis btaA gene, protein, L91 mutant.  
 DE Phospholipid synthesis; transgenic plant; fertilizer; enzyme;  
 XX betaine lipid; mutin.

XX Rhodobacter sphaeroideis.  
 OS Synthetic.

XX WO2005009115-A2.

XX 03-FEB-2005.

XX 15-JUL-2004; 2004WO-US022789.

XX 16-JUL-2003; 2003US-00620914.

XX (UNMS ) UNIV MICHIGAN STATE.

XX Benning C, Riekhof W;

XX WPI; 2005-112975/12.

XX New composition comprising a purified DNA having an oligonucleotide  
 PT sequence encoding a protein, useful in producing Betaine lipid compounds  
 PT e.g., Diacylglycerol-O-4'-(N,N,N-trimethyl) homoserine (DGTS).

XX Example 3; Page; 147pp; English.

XX The invention relates to a new composition comprising a purified DNA  
 CC having an oligonucleotide sequence appearing as ADW72747 (Chlamydomonas  
 CC reinhardtii Bta1 cDNA) or ADW72752 (Neurospora crassa Bta1 coding region,  
 CC encoding Bta1 proteins which are DGTS (diacylglycerol-O-4'-(N,N,N-trimethyl)homoserine) synthetic enzymes. Bta1 (and BtaA and BtaB) enzymes  
 CC allow the replacement of phospholipids with non-phosphorus containing  
 CC lipids in the cell membrane. A transgenic plant expressing the enzymes  
 CC would have a reduced need for phosphate-containing fertilizer. Also  
 CC included are an RNA transcribed from the purified DNA, antibodies  
 CC produced from the protein, a vector comprising the DNA or nucleic acid, a  
 CC host cell comprising the vector, transgenic plants comprising the vector,  
 CC and a protein translated from the RNA or encoded by the nucleic acid. The  
 CC composition is useful in producing Betaine lipid compounds e.g.,  
 CC Diacylglycerol-O-4'-(N,N,N-trimethyl) homoserine (DGTS). The composition  
 CC of the invention reduces the amount of phosphate fertilizer needed for  
 CC the optimal growth of crop plants. The present sequence represents a  
 CC mutated R. sphaeroideis BtaA or BtaB protein. NOTE: The present sequence  
 CC is not shown in the specification but was created by the indexer using  
 CC the wild-type protein and the information in example 3. Only one of the 3  
 CC BtaA mutants could be created, the other two (A201G and S399T) referred  
 CC to amino acids not present at those positions in the wild-type sequence.

XX Sequence 416 AA;

Query Match 11.0%; Score 381.5; DB 9; Length 416;  
 Best Local Similarity 28.1%; Pred. No. 5.5e-30;  
 Matches 127; Conservative 68; Mismatches 182; Indels 75; Gaps 20;  
 QY 231 TQGSIVYVWLRAPYVYVWLRGLPSVGHALH-----EERVERPPMPFPPTF---LVTQSW 280  
 Db 2 TQFALTHIP---APP---VAR---QIGAAVHRTSLLSAEGLMER--MPSRLFHGLVLPQIW 51  
 QY 281 EDPEPDMYMEINPKDVTVLTLTSGGCNVALNLLVQAGQVSVVDCNPAQSALLBLKKVAIQ 340  
 Db 52 EDPADVMAALAIRPGDRLVAIASGGCNVLSYLTQGGPSILAVDLSPAHRVALGRKLAAAR 111  
 QY 341 QL-EPEDVWQLFGEVHPRIEELYEKKLAPFLSQTSHNF-----SKRLWYFQHGGLYYQ 393  
 Db 112 TLPDHAADFLLFGRADLPQNAALYDRHIAIPALDGRSRRYWEARSPPGRRITQLFERGFYRH 171  
 QY 394 GGMKLCVWLQCLAVVLGLGKTVKRLANAPTMBEQRRLLWDSNNLIHFVKNQPKPLVWLFV 453  
 Db 172 GAGRFIGAHTLA--RAAGTDLRGFLDCPDTEAQRSFFVAHI-----GP-----LFE 217  
 QY 454 KFVSLVLFNKAVLWFGGVPKGQYALIKADG----IPIENYIARTMDGVAENSH-----503  
 Db 218 APVQALARRPAALFGLGIPPAQYALLAGDGDGVLPV-----LRQRLHLLCDF 267  
 QY 504 VRKQNYFYNYCLTGKFLRN---CPTYLREAPATILKSGVVDNLTVTSTFFMEELKARTY 560  
 Db 268 PLRENYFAFAIARRYPRFEGALPPYLEPTAFETLREN-AGRVQIENRSLTEALAAEPE 326  
 QY 561 TKV---ILMDHVDLMDMPVANELAECLAKQVAPGGIVWR---SASISPPYABLIQKAGF 614  
 Db 327 ESIHGFTLLDAQDMTDAQLTALWROVTRTAAPGARVIFRTGGADLLP--GRVPEILG 384  
 QY 615 DVCIRRTATQG--YMDRVNMYSSFYMARRKA 644  
 Db 385 HMRADRAAGQAGHAADRSALYGGFHLVRRDA 416

## RESULT 6

ABG72176  
 ID ABG72176 standard; protein; 416 AA.

XX AC ABG72176;

XX DT 05-FEB-2003 (first entry)

XX Rhodobacter sphaeroideis btaA protein.

XX Betaine lipid production; betaine lipid compound; DGTS;  
 KW diacylglycerol-O-4'-(N,N,N-trimethyl) homoserine; agricultural industry;  
 KW phosphate-containing fertilizer; btaA; enzyme.

XX Rhodobacter sphaeroideis.

XX Key Location/Qualifiers

XX Misc-difference 1 /note= "Encoded by GTC"

XX WO200283844-A2.

XX 24-OCT-2002.

XX 09-APR-2002; 2002WO-US011134.

XX 13-APR-2001; 2001US-0283812P.

XX 08-APR-2002; 2002US-00283812.

XX (UNMS ) UNIV MICHIGAN STATE.

XX Benning C, Riekhof W, Klug R;

XX WPI; 2003-058632/05.

XX N-PSDB; ABS58051.

XX

PT New composition comprising an isolated and purified DNA molecule, useful  
 PT for producing Betaine lipids, e.g. Diacylglycerol-1-O-4'-(N,N,N-trimethyl)  
 PT homoserine (DGTS) for agricultural applications.

XX Claim 11; Fig 9; 109pp; English.

PS The present invention relates to compositions and methods for producing  
 CC betaine lipids. The method comprises the expression of recombinant  
 CC enzymes (e.g. from Rhodospirillum rubrum) in host cells such as  
 CC bacteria, yeast, and plants to produce betaine lipids. The methods and  
 CC diacylglycerol-1-O-4'-(N,N,N-trimethyl) homoserine (DGTS). The methods and  
 CC compositions of the invention are useful in agricultural applications,  
 CC such that the amount of phosphate-containing fertilizer required for the  
 CC growth of a particular plant is decreased. The polynucleotide sequences  
 CC encoding the recombinant enzymes may be used to produce vectors which can  
 CC be used to produce transgenic plants. The present sequence represents R.  
 CC sphaeroides betA protein

XX Sequence 416 AA;

Query Match 11.0%; Score 379.5; DB 6; Length 416;

Best Local Similarity 28.1%; Pred. No. 8.9e-30;

Matches 127; Conservative 68; Mismatches 182; Indels 75; Gaps 20;

QY 231 TQGISPVVWLRAPYYVWIGRLPSVGHALH-----EERVERPPMPPTTF---LYTQSW 280

Db 2 TQFALTHLP---APP---VAR---QIGNAVHRTSLLSAGELMER--MFSRLFHGLVYPIW 51

QY 281 EDPEPMEVMEINPKDTVLTSTGGCNALNLVQAGQVVSVCNPAQSALLEKKVAIQ 340

Db 52 EDPAVDMALAIRPGDRLVAIASGCGNVLSYLTQGPISILAVDLSPAHVALGRKLAAAR 111

QY 341 QL-EFEDVWQFEGVHPRIEELKLAFLSQTSHNF-----SKRLWYFHGLYQ 393

Db 112 TLPDHAFFDLFGADLPFGNAALYDRHIAFDGRSRRYWEARSPPGRRIQLFERGYRH 171

QY 394 GGMGKLCWLQCLAVLGLKTKVKRLANAPTMBEQRLWDSNMLIHFVKNGKPLVWLFV 453

Db 172 GALTGFIGAHTLA--RAAGTDLRGFLDCPDIEAQRSFVYAH-----LFE 217

QY 454 KFSVLVFNKAVLWFGGVPGKQYALIKADG-----IPIENYIARTMDGVAENSH----- 503

Db 218 APVVOALARRPAALFGLGIPPAQYVALLAGDGDGLPV-----LRQLHRLLCDF 267

QY 504 VRQNYFYNNCLTGKFLRDN---CPTYLREAAATLKSQVVDNLVTSTNFMELKARTY 560

Db 268 PLRENYFAFOAIARIPRPGEGALPPYLEPTAFETLREN-AGRVQIENRSLTEALAAPE 326

QY 561 TKV---ILMDHVDWLDMPVANELAECLAKQVAPGIVWR---SASLSPPYAEILQKAGF 614

Db 327 ESTHGFTLLDAQMDMTAQITLWRQVTRTAAGARVIFRTGGAADLLP--GRVPEILG 384

QY 615 DVRCIRATQG---YMDRVNMYSSPYMARRKA 644

Db 385 HWRADRAAGAGHAADRSAYVGGPHLYRRDA 416

RESULT 7

ADW72706

ID ADW72706 standard; protein; 416 AA.

XX

AC ADW72706;

XX

DT 21-APR-2005 (first entry)

XX

XX Rhodospirillum rubrum betA gene, protein.

XX

XX Rhodospirillum rubrum betA gene, protein.

XX

XX Phospholipid synthesis; transgenic plant; fertilizer; enzyme;

XX

XX betaine lipid.

XX

XX Rhodospirillum rubrum betA gene, protein.

XX

XX Location/Qualifiers

FT Misc-difference 1 /note="Encoded by GTC"

XX

XX WO2005009115-A2.

XX

XX 03-FEB-2005.

XX

XX 15-JUL-2004; 2004WO-US022789.

XX

XX 16-JUL-2003; 2003US-00620914.

XX

XX (UNMS ) UNIV MICHIGAN STATE.

XX

XX Benning C, Riekhof W;

XX

XX WPI; 2005-112975/12.

XX

XX N-PSDB; ADW72704.

XX

XX New composition comprising a purified DNA having an oligonucleotide  
 PT sequence encoding a protein, useful in producing Betaine lipid compounds  
 PT e.g., Diacylglycerol-1-O-4'-(N,N,N-trimethyl) homoserine (DGTS).

PS Example 3; SEQ ID NO 3; 147pp; English.

XX The invention relates to a new composition comprising a purified DNA  
 CC having an oligonucleotide sequence appearing as ADW72747 (Chlamydomonas  
 CC reinhardtii BtAI cDNA) or ADW72752 (Neurospora crassa BtAI coding region,  
 CC encoding BtAI proteins which are DGTS (diacylglycerol-1-O-4'-(N,N,N-  
 CC trimethyl)homoserine) synthetic enzymes. BtAI (and BtAA and BtAB) enzymes  
 CC allow the replacement of phospholipids with non-phosphorus containing  
 CC lipids in the cell membrane. A transgenic plant expressing the enzymes  
 CC would have a reduced need for phosphate-containing fertilizer. Also  
 CC included are an RNA transcribed from the purified DNA, antibodies  
 CC produced from the protein, a vector comprising the DNA or nucleic acid, a  
 CC host cell comprising the vector, transgenic plants comprising the vector,  
 CC and a protein translated from the RNA or encoded by the nucleic acid. The  
 CC composition is useful in producing Betaine lipid compounds e.g.,  
 CC Diacylglycerol-1-O-4'-(N,N,N-trimethyl) homoserine (DGTS). The composition  
 CC of the invention reduces the amount of phosphate fertilizer needed for  
 CC the optimal growth of crop plants. The present sequence represents a  
 CC betaine lipid synthetic enzyme.

XX Sequence 416 AA;

Query Match 11.0%; Score 379.5; DB 9; Length 416;

Best Local Similarity 28.1%; Pred. No. 8.9e-30;

Matches 127; Conservative 68; Mismatches 182; Indels 75; Gaps 20;

QY 231 TQGISPVVWLRAPYYVWIGRLPSVGHALH-----EERVERPPMPPTTF---LYTQSW 280

Db 2 TQFALTHLP---APP---VAR---QIGNAVHRTSLLSAGELMER--MFSRLFHGLVYPIW 51

QY 281 EDPEPMEVMEINPKDTVLTSTGGCNALNLVQAGQVVSVCNPAQSALLEKKVAIQ 340

Db 52 EDPAVDMALAIRPGDRLVAIASGCGNVLSYLTQGPISILAVDLSPAHVALGRKLAAAR 111

QY 341 QL-EFEDVWQFEGVHPRIEELKLAFLSQTSHNF-----SKRLWYFHGLYQ 393

Db 112 TLPDHAFFDLFGADLPFGNAALYDRHIAFDGRSRRYWEARSPPGRRIQLFERGYRH 171

QY 394 GGMGKLCWLQCLAVLGLKTKVKRLANAPTMBEQRLWDSNMLIHFVKNGKPLVWLFV 453

Db 172 GALTGFIGAHTLA--RAAGTDLRGFLDCPDIEAQRSFVYAH-----LFE 217

QY 454 KFSVLVFNKAVLWFGGVPGKQYALIKADG-----IPIENYIARTMDGVAENSH----- 503

Db 218 APVVOALARRPAALFGLGIPPAQYVALLAGDGDGLPV-----LRQLHRLLCDF 267

QY 504 VRQNYFYNNCLTGKFLRDN---CPTYLREAAATLKSQVVDNLVTSTNFMELKARTY 560

Db 268 PLRENYFAFOAIARIPRPGEGALPPYLEPTAFETLREN-AGRVQIENRSLTEALAAPE 326

QY 561 TKV---ILMDHVDWLDMPVANELAECLAKQVAPGIVWR---SASLSPPYAEILQKAGF 614



```
Db 327 ESINGFTLLDAQDWMTDALQTLWRQVTRTAAPCARVIFRTGGAADLP--GRVPBEILG 384
QY 615 DVICIRATOG--YMDRVNMYSSFYMARRKA 644
Db 385 HWRADRAAGQAGHAADRSAYGGFHLVRRDA 416

RESULT 8
ABG72178
ID ABG72178 standard; protein; 416 AA.
XX
AC ABG72178;
DT
XX 05-FEB-2003 (first entry)
XX Agrobacterium tumefaciens btaA protein.
XX
XX Betaine lipid production; betaine lipid compound; DGTS;
KW diacylglycerol-0-4'-(N,N,N,-trimethyl) homoserine; agricultural industry;
KW phosphate-containing fertiliser; transgenic; btaA; enzyme.
XX
XX Agrobacterium tumefaciens.
XX
XX W0200283844-A2.
XX
XX 24-OCT-2002.
XX
XX 09-APR-2002; 2002WO-US011134.
XX
XX 13-APR-2001; 2001US-0283812P.
XX 08-APR-2002; 2002US-00283812.
XX
XX (UNMS ) UNIV MICHIGAN STATE.
XX
XX Benning C, Riekhof W, Klug R;
XX
XX WPI; 2003-058632/05.
XX N-PSDB; ABS58082.
XX
XX New composition comprising an isolated and purified DNA molecule, useful
PT for producing betaine lipids, e.g. Diacylglycerol-0-4'-(N,N,N,-trimethyl)
PT homoserine (DGTS) for agricultural applications.
XX
XX Claim 11; Fig 25; 109pp; English.
XX
XX The present invention relates to compositions and methods for producing
CC betaine lipids. The method comprises the expression of recombinant
CC enzymes (e.g. from Rhodospirillum rubrum) in host cells such as
CC bacteria, yeast, and plants to produce betaine lipid compounds including
CC diacylglycerol-0-4'-(N,N,N,-trimethyl) homoserine (DGTS). The methods and
CC compositions of the invention are useful in agricultural applications,
CC such that the amount of phosphate-containing fertilizer required for the
CC growth of a particular plant is decreased. The polynucleotide sequences
CC encoding the recombinant enzymes may be used to produce vectors which can
CC be used to produce transgenic plants. The present sequence represents the
CC R. sphaeroides btaA homologue from Agrobacterium tumefaciens
XX
XX Sequence 416 AA;

Query Match 9.4%; Score 326.5; DB 6; Length 416;
Best Local Similarity 27.8%; Pred. No. 2.9e-24;
Matches 111; Conservative 69; Mismatches 158; Indels 61; Gaps 16;

QY 275 LYTQSWEDPDPMEVMEINPKOTVLTGSGGNALMLVQAGGVSVSDCNPAQSALLEL 334
Db 44 VYPIWEDPDIEMAEELGEGHRIIVTGGGNNLAYSRLNPASIDVDLNPHEIALNKL 103
QY 335 KVAIQQL-EEFDVWQLFGGVHPIREELVEKLAFLPSLTSNFSK-----RLWVF 386
Db 104 KLAARHLPAAQDVRVHFGGRTSNGVDRFIAEHLDDATKAYWSKRTLSGRRISVF 163
QY 387 QHGLYYGGGKLCWVLQCLAVLGLGKTVKRLANAPTMBEQRLWDSNNLIHFVKNPK 446

Db 164 DRNIYRTGLGRFIGAGHIMARLHGVLKLT--EMAKTRTLDQEQFDSKVAPLF---DK 217
QY 447 PLV-WLFFKVSLSLVFNKAVLMFGGVPKQY---ALIKADGIPINNYIARTWDGVAENS 502
Db 218 PVVRWLTTRKSSL-----FGLGIPPRQVDELASLSDDG-TVASVLKERLEKLACNF 267
QY 503 HVRKONYFYNCITGKFLRDN---CPTYLREAFATLKSGVDNLTVSTNFFMEELK--- 556
Db 268 PL-SDNYFAWQAFARYPPEPHEGALPAYLKPEYKIRNNTA-RVAVHHATYTELLSRKP 325
QY 557 ARTYTKVILMDHVDLMDMPVANELAECLAKQVAPGGIVWRSAS-----LSPVPAEL 608
Db 326 ANGVDRIYILLDAQDWMTDVQLNELMSQISRTASGARVIFRTAARSKSVIEGRUSP----- 380
QY 609 IQKAGPDVR-----CIRRATQ-GYMDRVNMYSSFYMARR 641
Db 381 -----DIRNQWYLEERSNELNMDRSAYGGFHLVYQR 413

RESULT 9
ADW72732
ID ADW72732 standard; protein; 416 AA.
XX
AC ADW72732;
XX
XX 21-APR-2005 (first entry)
XX
XX Agrobacterium tumefaciens btaA gene, protein.
XX
XX Phospholipid synthesis; transgenic plant; fertilizer; enzyme;
KW betaine lipid.
XX
XX Agrobacterium tumefaciens.
XX
XX W02005009115-A2.
XX
XX 03-FEB-2005.
XX
XX 15-JUL-2004; 2004WO-US022789.
XX
XX 16-JUL-2003; 2003US-00620914.
XX
XX (UNMS ) UNIV MICHIGAN STATE.
XX
XX Benning C, Riekhof W;
XX
XX WPI; 2005-112975/12.
XX N-PSDB; ADW72731.
XX
XX New composition comprising a purified DNA having an oligonucleotide
PT sequence encoding a protein, useful in producing Betaine lipid compounds
PT e.g., Diacylglycerol-0-4'-(N,N,N,-trimethyl) homoserine (DGTS).
XX
XX Disclosure; SEQ ID NO 29; 147pp; English.
XX
XX The invention relates to a new composition comprising a purified DNA
CC having an oligonucleotide sequence appearing as ADW72747 (Chlamydomonas
CC reinhardtii Bta1 cDNA) or ADW72752 (Neurospora crassa Bta1 coding region,
CC encoding Bta1 proteins which are DGTS (diacylglycerol-0-4'-(N,N,N,-
CC trimethyl)homoserine) synthetic enzymes. Bta1 (and BtaA and BtaB) enzymes
CC allow the replacement of phospholipids with non-phosphorus containing
CC lipids in the cell membrane. A transgenic plant expressing the enzymes
CC would have a reduced need for phosphate-containing fertilizer. Also
CC included are an RNA transcribed from the purified DNA, antibodies
CC produced from the protein, a vector comprising the DNA or nucleic acid, a
CC host cell comprising the vector, transgenic plants comprising the vector,
CC and a protein translated from the RNA or encoded by the nucleic acid. The
CC composition is useful in producing Betaine lipid compounds e.g.,
CC Diacylglycerol-0-4'-(N,N,N,-trimethyl) homoserine (DGTS). The composition
CC of the invention reduces the amount of phosphate fertilizer needed for
CC the optimal growth of crop plants. The present sequence represents a
CC betaine lipid synthetic enzyme.
```

XX SQ Sequence 416 AA;

Query Match 9.4%; Score 326.5; DB 9; Length 416;  
 Best Local Similarity 27.8%; Pred. No. 2.9e-24;  
 Matches 111; Conservative 69; Mismatches 158; Indels 61; Gaps 16;

QY 275 LYTQSWEDPEPDMVEINPKDVTLTSTGGCNALMLLVQAGQVSVDCNPAQSALLEL 334  
 DB 44 VYQIWEDEPIDMEAMELGEGHRIIVTIGSGCNWLAIVLSRNPASIDVVDLPHHIALNKL 103

QY 335 KKVAIQOL-EFEDWQQLFGGVHPRIEELYEKKLAPLSOTSHNFSK-----RLWYF 386  
 DB 104 KLAAFRLHPAQDVRHFRAGRTSRNSGVYDFIAEHLDTATKAYNSKRTLSGRRRISVF 163

QY 387 QHGLYYQGGGKLCWLQCLAVVLGLGKTVKRLANAPTMEQRLWDSNMLIHVFNKGP 446  
 DB 164 DRNIYRTGLLGRFIGAGHIMARLHGKLT--EMAKRTLDEQRFDSKVAPLF----DK 217

QY 447 PLV-WLFPVKFVSLVLFNKAFLWFGGVPGKQY---ALIKADGIPNIENYIARTMDGVAENS 502  
 DB 218 PVRWLTYSKSSL-----FOLGIPPRQYDELASLSDDG-TVASVLKXERLEKACNF 267

QY 503 HVRKQNYFYNNCLTGKFLRDN---CPTYLREAAATLKGVDNLTVTSTNFFMEELK--- 556  
 DB 268 PL-SDNYFAWQAFARRYPEPEGALPAYLKPEYVEKIRNNTA-RVAVHHATYTELLSRKP 325

QY 557 ARTYTKVILMDHVDLMDPFAVLAELAECLAQVAPGGIVWRSAS-----LSPPYAEL 608  
 DB 326 ANGVDRIYLLDAQDWMVDVQLNELWSQISRTAASGARVIFRTAAEKSVIEGRILSP----- 380

QY 609 IQKAGFDVR-----CIRRTAQ-GYMDRVNMYSSFYMAR 641  
 DB 381 -----DIRNQWVLEERSNELNMDRSAYGGFHHIYQR 413

RESULT 10  
 ADW72744  
 ID ADW72744 standard; protein; 415 AA.  
 XX AC ADW72744;  
 XX DT 21-APR-2005 (first entry)  
 XX DE Mesorhizobium loti btaA gene, protein.  
 XX KW Phospholipid synthesis; transgenic plant; fertilizer; enzyme;  
 XX KW betaine lipid.  
 XX OS Mesorhizobium loti.  
 XX PN W02005009115-A2.  
 XX PD 03-FEB-2005.  
 XX PF 15-JUL-2004; 2004WO-US022789.  
 XX PR 16-JUL-2003; 2003US-00620914.  
 XX PA (UNMS) UNIV MICHIGAN STATE.  
 XX PI Benning C, Riekhof W;  
 XX DR WPI; 2005-112975/12.  
 XX DR N-PSDB; ADW72725.  
 XX XX  
 XX XX New composition comprising a purified DNA having an oligonucleotide  
 XX PT sequence encoding a protein, useful in producing Betaine lipid compounds  
 XX PT e.g., Diacylglycerol-0-4'-(N,N,N-trimethyl) homoserine (DGTS).  
 XX PS Disclosure; SEQ ID NO 41; 147pp; English.  
 XX PS The invention relates to a new composition comprising a purified DNA  
 XX CC

CC having an oligonucleotide sequence appearing as ADW72747 (Chlamydomonas  
 CC reinhardtii BtaI cDNA) or ADW72752 (Neurospora crassa BtaI coding region,  
 CC encoding BtaI proteins which are DGTS (diacylglycerol-0-4'-(N,N,N-  
 CC trimethyl)homoserine) synthetic enzymes. BtaI (and BtaA and BtaB) enzymes  
 CC allow the replacement of phospholipids with non-phosphorus containing  
 CC lipids in the cell membrane. A transgenic plant expressing the enzymes  
 CC would have a reduced need for phosphate-containing fertilizer. Also  
 CC produced are an RNA transcribed from the purified DNA, antibodies  
 CC host cell comprising the vector, transgenic plants comprising the vector,  
 CC and a protein translated from the RNA or encoded by the nucleic acid. The  
 CC composition is useful in producing betaine lipid compounds e.g.,  
 CC diacylglycerol-0-4'-(N,N,N-trimethyl) homoserine (DGTS). The composition  
 CC of the invention reduces the amount of phosphate fertilizer needed for  
 CC the optimal growth of crop plants. The present sequence represents a  
 CC betaine lipid synthetic enzyme.  
 XX SQ Sequence 415 AA;

Query Match 9.3%; Score 321; DB 9; Length 415;  
 Best Local Similarity 26.2%; Pred. No. 1.1e-23;  
 Matches 113; Conservative 70; Mismatches 176; Indels 72; Gaps 18;

QY 255 VGHALHEERV-----ERPPMPPTF---LYTQSWEDPEPDMVEINPKDVTLTSTG 304  
 DB 16 VGVKAVYQNRALS KAGISER--LFAFLFGLVLPQIWEDEPDMVEINPKDVTLTSTG 73

QY 305 GCNALMLLVQAGQVSVDCNPAQSALLELKKVAIQOLEFE-DVWQLFGGVHPRIEELY 363  
 DB 74 GCNLAFLYLRSPARIDAVDLNAAHIALNRKLEAVRRLPSQGLFRFFFGAATSHNSQAY 133

QY 364 EKKLAPFLSQTSHNFSKRLW-----YFQHGGLYYQGGGKLCWLQCLAVVLG----- 411  
 DB 134 DRFIAHLDPVSHYWERNRWGRRIAVDFRNFYQTGLLG-----LFIAMGHRTAK 185

QY 412 -LGKTVKRLANAPTMEEQRLWDSNMLIHVFNKGPPLVWLFVKFVSLVLFNKAFLWFGG 470  
 DB 186 FFGVNPAMWMEARNIGEQRRFFNEELAPVFDK---KLLKWATSRKASL-----FGL 233

QY 471 GVPKQY--ALIKADGIPNIENYIAR---TMDGVAENSHVRKQNYFYNNCLTGKFLRDN- 523  
 DB 234 GIPPAQYDSLITSGDGTMAVSKARLEKACDPFLEN-----NYFAWQAFARRYPNPG 287

QY 524 --CPTYLREAAATLKGVDNLTVTSTNFFMEEL---KARTYTKVILMDHVDLMDPFAV 578  
 DB 288 AALPAYLEKQNYETIR-GNIDRVAIHHAHLIEFLAGKDACTVDRFILLDAQDWMVDLQNL 346

QY 579 ELAECLAQVAPGGIVWRSAS---SLSP--PYAELIQKAGFDVRCIRRTATQGYMDRVNMY 633  
 DB 347 ALWSEISRTASAGARVIFRTAABPSLLPGRVSTSLDDQMDYQDEASREFSA--RDRSAIY 404

QY 634 SSFYMAR 644  
 DB 405 GGFHLYVKRTA 415

RESULT 11  
 ABG72180  
 ID ABG72180 standard; protein; 416 AA.  
 XX AC ABG72180;  
 XX DT 05-FEB-2003 (first entry)  
 XX DE Sinorhizobium meliloti btaA protein.  
 XX KW Betaine lipid production; betaine lipid compound; DGTS;  
 XX KW diacylglycerol-0-4'-(N,N,N-trimethyl) homoserine; agricultural industry;  
 XX KW phosphate-containing fertilizer; transgenic; btaA; enzyme.  
 XX OS Sinorhizobium meliloti.  
 XX PS W0200283844-A2.  
 XX CC

XX PD 24-OCT-2002.  
 XX XX 09-APR-2002; 2002WO-US011134.  
 XX XX 13-APR-2001; 2001US-0283812P.  
 XX XX 08-APR-2002; 2002US-00283812.  
 XX PA (UNMS ) UNIV MICHIGAN STATE.  
 XX XX Benning C, Riekhof W, Klug R;  
 XX XX WPI; 2003-058632/05.  
 DR N-PSDB; ABS58084.  
 XX XX New composition comprising an isolated and purified DNA molecule, useful  
 PT for producing Betaine lipids, e.g. Diacylglycerol-0-4'-(N,N,N-trimethyl)  
 PT homoserine (DGTS) for agricultural applications.  
 XX PS Claim 11; Fig 29; 109pp; English.  
 XX CC The present invention relates to compositions and methods for producing  
 CC betaine lipids. The method comprises the expression of recombinant  
 CC enzymes (e.g. from Rhodospirillum rubrum) in host cells such as  
 CC bacteria, yeast, and plants to produce betaine lipid compounds including  
 CC diacylglycerol-0-4'-(N,N,N-trimethyl) homoserine (DGTS). The methods and  
 CC compositions of the invention are useful in agricultural applications,  
 CC such that the amount of phosphate-containing fertilizer required for the  
 CC growth of a particular plant is decreased. The polynucleotide sequences  
 CC encoding the recombinant enzymes may be used to produce vectors which can  
 CC be used to produce transgenic plants. The present sequence represents the  
 CC R. rubrum betaine synthase gene from Rhodospirillum rubrum  
 XX SQ Sequence 416 AA;  
 Query Match 8.4%; Score 292.5; DB 6; Length 416;  
 Best Local Similarity 26.0%; Pred. No. 1e-20;  
 Matches 102; Conservative 71; Mismatches 174; Indels 45; Gaps 16;  
 QY 275 LYTQSWEDPDPMEVMEINPKDVTLTGSGCNALNLLVQAGQVSVDCNPAQSALLEL 334  
 DB 45 VYQIWEEDPVDMEAMQIRPHRIVTIGSGCNMLTYLSAEPARIDVDLNPHEALNRL 104  
 QY 335 KKVAIQOL-BFEDVWOLFQ-EGVHPRIEYKELAPFLSQTSHNFWMS-----KRLWY 385  
 DB 105 KLSAFRHLPSHKDVRFLAVEGRTN-GQAYDVFLAPKLPATRAYWNGRDLTGRRRIGV 163  
 QY 386 FQHGLYYQGGMGKLCWVLCQAVLGLGKTVKRLANAPTMEORRLWDSNMLHFVNKGP 445  
 DB 164 FGRNVYRTGLGRFISASHALRLHGINP--EDFVKARSMEQRFDDKLAFLP---E 217  
 QY 446 KPLV-WLFVKFVSLVLFNKAFLWFGGVPGKQYALIKADGIPYENIARTMDGVAENSHV 504  
 DB 218 RPVIRWITSKSSU-----FGLGIPQQDELAS--LSREKSVAAVLRNLEKLTCTC 266  
 QY 505 R---KONYFYNNCLTKGFLRDN---CPTYLREAAFPATLKSQVVDNLVTSTNFFMEEL--- 555  
 DB 267 HPFLRDNYFAWQAPARYPRPDGELPPYLOASRYEARDN-AERVEVHHSATTELLAGK 325  
 QY 556 KARTYTKVILMDHVDLMDVANELAECLAKQVAPGGIVWRS---ASLSPP--YAEILQ 610  
 DB 326 PAASVDVRYLLDAQDMMWTDQDLNLMTEITRTADAGAWIFRTAAEASILPGRSLSTLLD 385  
 QY 611 KAGFDVRCIRATQGYMDRVNMTSSSFYMAR 642  
 DB 386 QWTYDAETSURL--GAEDRSATYGGFHYRKK 415  
 RESULT 12  
 ADW72736  
 ID ADW72736 standard; protein; 416 AA.  
 XX AC ADW72736;

XX DT 21-APR-2005 (first entry)  
 XX DE Sinorhizobium meliloti btaA gene, protein.  
 XX DE Phospholipid synthesis; transgenic plant; fertilizer; enzyme;  
 XX DE betaine lipid.  
 XX OS Sinorhizobium meliloti.  
 XX PN WO2005009115-A2.  
 XX PD 03-FEB-2005.  
 XX PF 15-JUL-2004; 2004WO-US022789.  
 XX PR 16-JUL-2003; 2003US-00620914.  
 XX PA (UNMS ) UNIV MICHIGAN STATE.  
 XX PI Benning C, Riekhof W;  
 XX WPI; 2005-112975/12.  
 DR N-PSDB; ADW72735.  
 XX PT New composition comprising a purified DNA having an oligonucleotide  
 PT sequence encoding a protein, useful in producing Betaine lipid compounds  
 PT e.g., Diacylglycerol-0-4'-(N,N,N-trimethyl) homoserine (DGTS).  
 XX PS Disclosure; SEQ ID NO 33; 147pp; English.  
 XX CC The invention relates to a new composition comprising a purified DNA  
 CC having an oligonucleotide sequence appearing as ADW72747 (Chlamydomonas  
 CC reinhardtii Bta1 cDNA) or ADW72752 (Neurospora crassa Bta1 coding region,  
 CC encoding Bta1 proteins which are DGTS (diacylglycerol-0-4'-(N,N,N-  
 CC trimethyl)homoserine) synthetic enzymes. Bta1 (and BtaA and BtaB) enzymes  
 CC allow the replacement of phospholipids with non-phosphorus containing  
 CC lipids in the cell membrane. A transgenic plant expressing the enzymes  
 CC would have a reduced need for phosphate-containing fertilizer. Also  
 CC included are an RNA transcribed from the purified DNA, antibodies  
 CC produced from the protein, a vector comprising the DNA or nucleic acid, a  
 CC host cell comprising the vector, transgenic plants comprising the vector,  
 CC and a protein translated from the RNA or encoded by the nucleic acid. The  
 CC composition is useful in producing Betaine lipid compounds e.g.,  
 CC Diacylglycerol-0-4'-(N,N,N-trimethyl) homoserine (DGTS). The composition  
 CC of the invention reduces the amount of phosphate fertilizer needed for  
 CC the optimal growth of crop plants. The present sequence represents a  
 CC betaine lipid synthetic enzyme.  
 XX SQ Sequence 416 AA;  
 Query Match 8.4%; Score 292.5; DB 9; Length 416;  
 Best Local Similarity 26.0%; Pred. No. 1e-20;  
 Matches 102; Conservative 71; Mismatches 174; Indels 45; Gaps 16;  
 QY 275 LYTQSWEDPDPMEVMEINPKDVTLTGSGCNALNLLVQAGQVSVDCNPAQSALLEL 334  
 DB 45 VYQIWEEDPVDMEAMQIRPHRIVTIGSGCNMLTYLSAEPARIDVDLNPHEALNRL 104  
 QY 335 KKVAIQOL-BFEDVWOLFQ-EGVHPRIEYKELAPFLSQTSHNFWMS-----KRLWY 385  
 DB 105 KLSAFRHLPSHKDVRFLAVEGRTN-GQAYDVFLAPKLPATRAYWNGRDLTGRRRIGV 163  
 QY 386 FQHGLYYQGGMGKLCWVLCQAVLGLGKTVKRLANAPTMEORRLWDSNMLHFVNKGP 445  
 DB 164 FGRNVYRTGLGRFISASHALRLHGINP--EDFVKARSMEQRFDDKLAFLP---E 217  
 QY 446 KPLV-WLFVKFVSLVLFNKAFLWFGGVPGKQYALIKADGIPYENIARTMDGVAENSHV 504  
 DB 218 RPVIRWITSKSSU-----FGLGIPQQDELAS--LSREKSVAAVLRNLEKLTCTC 266  
 QY 505 R---KONYFYNNCLTKGFLRDN---CPTYLREAAFPATLKSQVVDNLVTSTNFFMEEL--- 555





**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: March 14, 2006, 01:46:11 ; Search time 32.3 Seconds

(without alignments)

5037.237 Million cell updates/sec

Title: US-10-620-914-44

Perfect score: 3619

Sequence: 1 atggggctgggtcgtgcagg.....gcgcgaagaaggacaactaa 1947

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 3735138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlh  
-Q=/abs/ABSSWEB pool/US10620914/runat\_13032006\_101936\_5626/app query.fasta\_1  
-DB=Published Applications AA Main -QFMT=fastan -SUFFIX=n2p.rapbm  
-MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1  
-MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DLOCALIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptco -NORM=ext  
-HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abs05h  
-USER=US10620914 @CGN 1 1 134 @runat\_13032006\_101936\_5626 -NCFU=6 -ICPU=3  
-NO\_MMAP -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DRV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA Main:

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	3463	95.7	648	4	US-10-620-914-45
2	891	24.6	752	5	Sequence 45, Appl
3	850.5	23.5	908	4	Sequence 7021, Ap
C 4	476.5	13.2	19662	4	US-10-620-914-50
5	459	12.7	19723	4	Sequence 6, Appl
C 6	457.5	12.6	19723	4	Sequence 5, Appl
C 7	456.5	12.6	19658	4	US-10-084-846A-5
8	446	12.3	19608	4	US-10-084-846A-7
9	443.5	12.3	19695	4	US-10-084-846A-3
C 10	436	12.0	19608	4	US-10-084-846A-8
C 11	424	11.7	19725	4	US-10-084-846A-4

12	407	11.2	19725	4	US-10-084-846A-4	Sequence 4, Appl
C 13	383.5	10.6	19695	4	US-10-084-846A-3	Sequence 3, Appl
14	379.5	10.5	416	4	US-10-118-495-3	Sequence 3, Appl
15	379.5	10.5	416	4	US-10-620-914-3	Sequence 3, Appl
C 16	370	10.2	900	5	US-10-450-763-33892	Sequence 33892, A
17	366.5	10.1	1084	5	US-10-794-514A-336	Sequence 336, App
C 18	357	9.9	1084	5	US-10-794-514A-336	Sequence 336, App
C 19	352	9.7	599	5	US-10-450-763-33889	Sequence 33889, A
20	342.5	9.5	19662	4	US-10-084-846A-6	Sequence 6, Appl
21	341.5	9.4	437	4	US-10-437-963-132079	Sequence 132079,
22	335.5	9.3	465	4	US-10-437-963-174113	Sequence 174113,
23	326.5	9.0	416	4	US-10-118-495-29	Sequence 29, Appl
24	326.5	9.0	416	4	US-10-620-914-29	Sequence 29, Appl
25	321	8.9	415	4	US-10-118-495-41	Sequence 41, Appl
26	321	8.9	415	4	US-10-620-914-41	Sequence 41, Appl
27	320.5	8.9	19652	4	US-10-084-846A-7	Sequence 7, Appl
28	308	8.5	455	4	US-10-437-963-161457	Sequence 161457,
C 29	307	8.5	497	4	US-10-437-963-125004	Sequence 125004,
30	304	8.4	599	5	US-10-450-763-33889	Sequence 33889, A
31	303.5	8.4	4440	4	US-10-052-586-525	Sequence 525, App
32	303.5	8.4	4440	4	US-10-174-590-525	Sequence 525, App
33	303.5	8.4	4440	4	US-10-176-758-525	Sequence 525, App
34	303.5	8.4	4440	4	US-10-175-737-525	Sequence 525, App
35	303.5	8.4	4440	4	US-10-174-581-525	Sequence 525, App
36	303.5	8.4	4440	4	US-10-176-483-525	Sequence 525, App
37	303.5	8.4	4440	4	US-10-176-749-525	Sequence 525, App
38	303.5	8.4	4440	4	US-10-176-914-525	Sequence 525, App
39	303.5	8.4	4440	4	US-10-176-915-525	Sequence 525, App
40	303.5	8.4	4440	4	US-10-173-706-525	Sequence 525, App
41	303.5	8.4	4440	4	US-10-175-738-525	Sequence 525, App
42	303.5	8.4	4440	4	US-10-175-752-525	Sequence 525, App
43	303.5	8.4	4440	4	US-10-176-482-525	Sequence 525, App
44	303.5	8.4	4440	4	US-10-176-757-525	Sequence 525, App
45	303.5	8.4	4440	4	US-10-176-913-525	Sequence 525, App

#### ALIGNMENTS

##### RESULT 1

US-10-620-914-45  
; Sequence 45, Application US/10620914  
; Publication No. US20040093639A1  
; GENERAL INFORMATION:  
; APPLICANT: Benning, Christoph  
; APPLICANT: Riekhof, Wayne  
; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids  
; FILE REFERENCE: MSU-07769  
; CURRENT APPLICATION NUMBER: US/10/620,914  
; CURRENT FILING DATE: 2003-07-16  
; PRIOR APPLICATION NUMBER: 10/118,495  
; PRIOR FILING DATE: 2002-04-08  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 45  
; LENGTH: 648  
; TYPE: PRT  
; ORGANISM: Chlamydomonas reinhardtii  
US-10-620-914-45

Alignment Scores:  
Pred. No.: 1.73e-204 Length: 648  
Score: 3463.00 Matches: 648  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 95.7% Indels: 0  
DB: 4 Gaps: 0

US-10-620-914-44 (1-1947) x US-10-620-914-45 (1-648)

Qy 1 ATGGGGTCTGGTCTGACGCGCGCGCTGACACCAAGAGACTTCTCCCTCGAG 60  
|||||  
Db 1 MetGlySerGlyArgAspGlyArgProAlaSerTyrThrIlyAsnPhenSerLeuGlu 20

QY 61 AAGCTCAAGCTCAGAGCATGAGATGACCTGACCGTTCTGCGCCATATGTGTTGGC 120  
DB 21 LysLeuLysLeuSerSerMetLysAspAspLeuThrValLeuArgHisMetTrpPheGly 40  
QY 121 AGCAAGAGGCGGATGATCAGCTGCTCGCTCGAGAGCTTCTACGGGCGCCAGCGCGCT 180  
DB 41 SerLysLysGlyAspAspHisAlaAlaArgLeuGluSerPheTyrGlyProGlnAlaAla 60  
QY 181 GCCTTTGCTGCGCGCTGCGGAGCGCTCGAACCTCATCTCGGTTGACCTGGTGGTGGC 240  
DB 61 AlaPheAlaAlaArgLeuAlaGluArgSerAsnLeuIleTrpValAspLeuGlyGly 80  
QY 241 ACTGGGAGATGTCGATATGATGCTGATTACATCGACCTGGCGAAGTTCAAGTCCATC 300  
DB 81 ThrGlyGluAenValAspMetMetAlaAspTyrIleAspLeuAlaLysPheLysSerIle 100  
QY 301 TACGTGGTCGACCTGTGCACTCGCTGTGAGGTGGCCAAAGAGAGAGGCGCAAG 360  
DB 101 TyrValValAspLeuLysHisSerLeuLysGluValAlaLysLysAlaLysAlaLys 120  
QY 361 GCGTGGAAAGATGTCAGGTGTCGAGGCGGACGCTTGCCAAATTTGCGCCCTCGAGGCG 420  
DB 121 GlyTrpLysAsnValGlnValGluAlaAspAlaCysGlnPheAlaProGluGly 140  
QY 421 ACCGGAGCGCTCATCACCCTTCTCTACCTGCTCAGCATGATCCACCGTTCCACACGTC 480  
DB 141 ThrAlaThrLeuIleThrPheSerTyrSerLeuThrMetIleProProPheHisAsnVal 160  
QY 481 ATCGACAGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540  
DB 161 IleAspGlnAlaCysSerTyrLeuSerGlnAspGlyLeuValGlyValAlaAspPheTyr 180  
QY 541 GTGAGCGGCAAGTACGACCTGCGCTGCGCAGATGCGCTGCTGCTGCTGCTGCTGCTGCTG 600  
DB 181 ValSerGlyLysTyrAspLeuProLeuArgGlnMetProTrpSerArgArgPhePheTrp 200  
QY 601 CGATCGATCTTCGACATCGACAACTGACATCGGCGCGCGCGCGCGCTGCTGAG 660  
DB 201 ArgSerIlePheAspIleAspAsnIleAspIleGlyProGluArgAlaTyrLeuGlu 220  
QY 661 CAGAGCTGGAGCGGTGGGAGCAGAACCCAGGCTGATCCCTGCTGCTGCTGCTGCTGCTG 720  
DB 221 GlnLysLeuGluArgValTrpGluGlnAsnThrGlnGlySerIleProTyrValProTrp 240  
QY 721 CTGCGCGCGCCCTACTAGCTGTGATGTCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGAC 780  
DB 241 LeuArgAlaProTyrTyrValTrpIleGlyArgLeuProSerValGlyHisAlaLeuHis 260  
QY 781 GAGGAGCGGTGGAGCG 840  
DB 261 GluGluArgValGluArgProMetPheProProMetPheProProPheLeuTyrThrGlnSerTrp 280  
QY 841 GAGGACCGCGAGCG 900  
DB 281 GluAspProGluProAspMetGluValMetGluIleAsnProLysAspThrValLeuThr 300  
QY 901 CTGACTAGCGCGCTGCAATCCCTGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960  
DB 301 LeuThrSerGlyGlyCysAsnAlaLeuAsnLeuValGlnGlyAlaGlyGlnValVal 320  
QY 961 TCGGTGAGCTGCAACCCCGCGAGTCGCGGCTTCTGGAGCTGAAGAGTGGCCATTGAG 1020  
DB 321 SerValAspCysAsnProAlaGlnSerAlaLeuLeuGluLeuLysLysValAlaIleGln 340  
QY 1021 CAGCTGGAGTTGAGAGCTGGAGCTGTCGCGAGGCGGTGCGCGCGCGCGCGCGCGCGCG 1080  
DB 341 GlnLeuGluPheGluAspValTrpGlnLeuPheGlyGlyValHisProArgIleGlu 360  
QY 1081 GAGCTGTACGAGAAGCTGGCGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140  
DB 361 GluLeuTyrGluLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrpSer 380

QY 1141 AAGCGCTCTGGTACTTCCAGCACGGCTGTACTACAGGGCGGCATGGGCAAGCTGTGC 1200  
DB 381 LysArgLeuTrpTyrPheGlnHisGlyLeuTyrTyrGlnGlyMetGlyLysLeuCys 400  
QY 1201 TGGGTGTGAGTCCCTGGCGCTGGTCTGGAGCTGGGCAAGACCGTCAAGCGCTCGCC 1260  
DB 401 TrpValLeuGlnCysLeuAlaValValLeuGlyLeuGlyLysThrValLysArgLeuAla 420  
QY 1261 AACGCGCCCAATGGAGGAGCGCGCTGTGGGACAGCAACATGCTCATCCACTTC 1320  
DB 421 AsnAlaProThrMetGluGlnArgLeuTrpAspSerAsnMetLeuIleHisPhe 440  
QY 1321 GTGAAGAACCGGCGCAAGCGCTGTGGTCTGTCTCAAGTTCTGAGCGCTCGTCTC 1380  
DB 441 ValLysAsnGlyProLysProLeuValTrpLeuPheValLysPheValSerLeuValLeu 460  
QY 1381 TTCAACAGCGCGTGTGTGCTGGCGGCGCTGCGGCAAGCAAGTACCGCTGATC 1440  
DB 461 PheAsnLysAlaValLeuTrpPheGlyGlyValProGlyLysGlnTyrAlaLeuIle 480  
QY 1441 AAGCGGACCGCATCCCATTTGAGAACTGACATCGCGCGCACCATGGAGCGGCTGCGGAG 1500  
DB 481 LysAlaAspGlyIleProIleGluAsnTyrIleAlaArgThrMetAspGlyValAlaGlu 500  
QY 1501 AACTCGCAGCTGCGCAAGCAGAACTACTTCTACTACAACTGCTCTCACCGGCAAGTTCCTG 1560  
DB 501 AsnSerHisValArgLysGlnAsnTyrPheTyrTyrAsnCysLeuThrGlyLysPheLeu 520  
QY 1561 CGCGCAACTGCGCCCACTACCTGCGGAGCGGCTTGGCCACCTCAGAGTGGCGTG 1620  
DB 521 ArgAspAsnCysProThrTyrLeuArgGluAlaAlaPheAlaThrLeuLysSerGlyVal 540  
QY 1621 GTGACAACTGACCGCTCCACCACCTTCTCATGAGGAGCTCAAAGCGCGCACCTTAC 1680  
DB 541 ValAspAsnLeuThrValSerThrAsnPhePheMetGluGluLeuLysAlaArgThrTyr 560  
QY 1681 ACCAAGTGATTTCTGATGGACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740  
DB 561 ThrLysValIleLeuMetAspHisValAspTrpLeuAspMetProValAlaAsnGluLeu 580  
QY 1741 GCGAGTGCCTTGGCCAAAGAGTTGCGCGCGCGCGCATGCTGCTGCTGCTGCTGCTGCTG 1800  
DB 581 AlaGluCysLeuAlaLysGlnValAlaProGlyGlyIleValIleTrpArgSerAlaSer 600  
QY 1801 CTCAGCGCGCTTACGCGAGCTGCTCAGAGGCGGCTGCTGCTGCTGCTGCTGCTGCTG 1860  
DB 601 LeuSerProProTyrAlaGluLeuIleGlnLysAlaGlyPheAspValArgCysIleArg 620  
QY 1861 CGCGCACTCAGGCTTACATGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1920  
DB 621 ArgAlaThrGlnGlyTyrMetAspArgValAsnMetTyrSerSerPheTyrMetAlaArg 640  
QY 1921 CGGAGGCGCGCAAGAGGACAAC 1944  
DB 641 ArgLysGlyAlaLysLysAspAsn 648

## RESULT 2

US-10-741-849-7021  
; Sequence 7021, Application US/10741849  
; Publication No. US20050019931A1  
; GENERAL INFORMATION:  
; APPLICANT: Roemer, Terry  
; APPLICANT: Jiang, Bo  
; APPLICANT: Boone, Charles  
; APPLICANT: Busey, Howard  
; TITLE OF INVENTION: Nucleic Acids Encoding Anti-fungal Drug Targets and Methods of  
; FILE REFERENCE: 10182-023-999  
; CURRENT APPLICATION NUMBER: US/10/741,849  
; CURRENT FILING DATE: 2003-12-19  
; PRIOR APPLICATION NUMBER: US 60/434,832  
; PRIOR FILING DATE: 2002-12-19  
; NUMBER OF SEQ ID NOS: 8000

SOFTWARE: PatentIn version 3.2  
 SEQ ID NO 7021  
 LENGTH: 752  
 TYPE: PRT  
 ORGANISM: Candida albicans  
 US-10-741-849-7021

## Alignment Scores:

Pred. No.: 2,69e-46 Length: 752  
 Score: 891.00 Matches: 229  
 Percent Similarity: 46.6% Conservative: 98  
 Best Local Similarity: 32.7% Mismatches: 232  
 Query Match: 24.6% Indels: 142  
 Gaps: 21

US-10-620-914-44 (1-1947) x US-10-741-849-7021 (1-752)

```

QY 151 CTGGAGACTTCTACGGCCCGCCAGCGCTGCCCTT----- 186
Db 70 LeuGluSerPheTyrLysAsnGlnAlaHisIleTyrAspAsnThrArgGluPheLeuLeu 89
QY 187 -----GTCGCCCGCTGGCGAGCGCTCGAACCTC 216
Db 90 LysGlyArgGlnGluCysLeuArgLeuAlaIleSerHisLeuProLysLysLysAspLeu 109
QY 217 ATCTGGTTGACCTGGGTGGGCTGCTGGGAGAAATGCGATATGATGGCTGATTACATC 276
Db 110 IleTrpIleAspIleGlyGlyThrGlySerAsnIleGluPheMetAspGluIleSer 129
QY 277 GACCTGGCGAG---TTCAAGTCCATCTACCTGCTGACCTGCTGCCACTCGCTGGAG 333
Db 130 LysIleSerGluAsnPheLysAlaValTyrLeuValAspLeuSerProSerLeuCysGlu 149
QY 334 GTGGCCAAAGAAAGCGCAAGCGCTGGAAGATGTCAGGTGCTGGAGCGCGAC 393
Db 150 ValAlaLysAlaArgPheGluAlaHisGluTrpThrAsnValHisValLeuValAlaAsp 169
QY 394 GCTTGCCAAATTTGGCCCTGGAGGGCAACCGGAGCTCATCATCTTCTCTACTCGCTC 453
Db 170 AlaCysAspPheThrIleAspTyrAspSerAlaAspLeuIleThrPheSerTyrSerLeu 189
QY 454 ACGATGATTCACCGTCCCAACGTCATCACCAGCGTCTGCTGCTACCTGCTCCCAAGAC 513
Db 190 SerMetIleProThrPheAsnAlaAlaIleAspAsnAlaValSerLysLeuAspMetGlu 209
QY 514 GGCCTGGTGGCGTTCGCGACTTCTACGTGAGC-----GGCAAGTAC 555
Db 210 GlyIleIleAlaThrValAspPheGlyIleGlnSerSerAspThrSerMetGlyArgIle 229
QY 556 GAC-----CTGCCCTCGCCAGATGCCGTGCTGGTCCGCCGCTTCTCTGGCGA 603
Db 230 AsnThrValGlyLysLeuValAsnArgAspIleProTrpIleLeuArgAsnPheTrpArg 249
QY 604 TCGATCTTCGACATCGACAATGACATCGCCCGCCGAGCGCGCTACCTACCTGGAGAG 663
Db 250 IleTrpPheGluAlaAspLysValPheLeuAspSerSerArgArgAsnTyrLeuGluTyr 269
QY 664 AAGCTGGAGCGGTGGGAGCAGAACACCCAG-----GGTTCGATCCCTAC 711
Db 270 LysPheGlyThrValLysSerLeuAsnSerTyrAsnLysAlaLeuGlyLysIle----- 287
QY 712 GTGCCGTGGTGGCGCCCTACTAGTGTGGATGGCCCGCTGCCAGCGTTGGCCAC 771
Db 288 -----ProTyrTrpIleTrpIleGlyCysAspLysSerLysSerHis 301
QY 772 GCCCTGCACGAGAG-----CGCGTGGAGCGCGCCCGCATGTTCCCGCCAC 819
Db 302 ThrIleLeuGluArgLeuAsnCysLeuAlaThrGluSerProTyrLeuAlaProThrThr 321
QY 819 ----- 819
Db 322 ThrProIleAlaAsnGlnLeuGluAspIleProIleSerLysGlyHisGluAlaLeu 341

```

```

QY 819 ----- 819
Db 342 IleAsnLeuGlnLysAsnLeuProTyrProSerMetTyrTyrGlnLysGluTyrTrpArg 361
QY 820 -----TTCTGTGACAG-831
Db 362 ValTyrTyrAspGluMetAsnProLeuTyrGluGlnPheLysAsnGlnTrpIleTyrAla 381
QY 832 CAGTCGTGGGAGGACCCCGAGCGATATGAGGTGATGGAGATCAACCCCAAGACACG 891
Db 382 PheThrTrpGluAspProArgGluAspHisLysLeuLeuAsnPheThrSerAspAspThr 401
QY 892 GTGCTGACCCCTGACTAGCGCGGCTGCAATCCCTGAACTG-----CTGGTGACGGG 945
Db 402 ValLeuAlaIleThrSerAlaGlyAspAsnIleLeuSerTyrAlaSerLeuProThrPro 421
QY 946 GCGCGCAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1005
Db 422 ProLysLysIleHisAlaValAspLeuAsnProCysGlnAsnHisLeuLeuGluLeuLys 441
QY 1006 AAGTGGCCCATTCACGAGCTGGAGCTGGAGACGTGGCGAGCTGTTCCGGCAGGCGGTG 1065
Db 442 LeuAlaSerPheArgCysLeuSerGlnGluGlnIleTrpSerMetPheGlyGluGlyLys 461
QY 1066 CACCCCGCATTTAGGAGCTGTAGAGAAGCTGGCGCCCTTCCTGTCGCAACACGAC 1125
Db 462 IleGluAsnPheAsnAspLeuLeuIleAspThrLeuAlaProHisMetSerSerAsnAla 481
QY 1126 CACAACCTTCTGGTCCAGCGC-----CTCTGGTACTTCCAGCAGCGCTCTACTACCAG 1179
Db 482 PheGlnTyrTrpMetAspLysGlyProLysThrPheSerGlyLysGlyLeuTyr----- 499
QY 1180 GCGCGCATGGCGAGCTGTGCTGGTGTGTCAGTGC-----CTGGCGCTGGTGTGCTG 1230
Db 500 ---AspThrGlyPheSerArgTrpAlaLeuArgLeuSerArgTyrValPheLysValCys 518
QY 1231 GGACTGGGCAAGACCGCTCAAGCGCTCGCCCAACCGCGCCCAATGGAGGAGAGCGCGCT 1290
Db 519 GlyValSerLysTyrValGluGluCysAlaAlaThrThrMetGluGluGlnLeuArg 538
QY 1291 CTGTGGGACACACATGCTCATCCATCTCGTGAAGACGGCGCCCAAGCGCTGTGTGG 1350
Db 539 IleTrpAsnGluHisLeu-----LysProThr----- 547
QY 1351 CTGTTCTCTCAAGTTCTGTG---AGCTGTGTGCTTTCACAGGCGCTGCTGTGTGTGGC 1407
Db 548 LeuPheAsnProValValGlySerLeuLeuValGlyAsnProMetPheLeuTrpLysAla 567
QY 1408 GCGCGCTGCGGCAAGCAGTACGCTGATCAAGCGGAGCGGATCCCCCATTTGAGAAC 1467
Db 568 LeuGlyValProAlaAsnGlnAlaLeuMet-----GlyProSerValIleLys 584
QY 1468 TACATCGCGCGCACCATTGGAGCGGTGGCGGAGAACTCGCACGTCGCCAAGACAGACTAC 1527
Db 585 TyrValValAspThrLeuAspProIleLysArgSerMetIleSerAsnAspAsnTyr 604
QY 1528 TTCTACTACAACTCCCTCAGCGCAAGTCTCGCGCACAACTGCCCGCCACTCTGTGGC 1587
Db 605 PheTyrTyrLeuCysMetMetGlyArgTyrThrLysAsnAsnCysProAspTyrLeuThr 624
QY 1588 GAGCGCGCTTCGCCACCCCTCAAG-----AGTGGCGTGTG 1623
Db 625 ThrLysGlyPheAsnArgLeuSerSerThrAlaAlaThrAlaSerGlySerSerProIle 644
QY 1624 GACAACTGACCGCTCTCCACCAACTTCTTCATGAG-----GAGCTCAAGCGCGC 1674
Db 645 AspAsnLeuArgIleHisThrAspThrLeuAsnGluValPheGlyArgLeuLysGluLys 664
QY 1675 ACCTACACCAAGGTGATTCGTGAGCACCATGCTGCTGGCTGGTGGAT-----ATGCC 1725
Db 665 SerIleThrIleAlaIleIleMetAspHisMetAspTrpPheAspProAsnGlyArgAsp 684
QY 1726 GTGGCCAAACGAGCTGGCGGAGTGCCTGCGCCCAAGCAGGTGTGCGCGCGGCGCATC 1785

```



```

QY 1351 CTGTCGTCACGTCGTCGAGC---CTGGTGTCTTCAACAAGCGCGTGTGTGGTTCGGC 1407
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 667 LeuLeuAsnArgPheValAsnGlyLeuValLeuSerSerAspAlaPheLeuTrpSerAla 686
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1408 GCGCGCGTGGCGGCAAGCAGTACGCGCTGATCAAGCGCGAC-----1449
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 687 LeuGlyValProLeuAsnGlnValAlaMetIleGluAlaAspTyrHisArgArgSerIle 706
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1450 -----GGCATCCCATTCAGAAC 1467
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 707 SerSerSerThrThrProSerSerLysGluLysProSerArgAlaGluAlaIleLeuHis 726
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1468 TACATCGCGGCACCATGACCGCGTGGCGGAGAACTGCAGCTGGCGCAAGCAAGTAC 1527
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 727 TyrThrThrSerThrLeuAspProValLeuSerThrSerHisLeuAlaSerAspAsnPro 746
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1528 TTCTACTCAACTGCTCAGCGCAAGTTCCTCGCGCAACTGCGCCACCTACTCTGCGC 1587
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 747 TyrTyrLeuValCysValLeuGlyGlnTyrThrArgGlnCysHisProAspTyrLeuSer 766
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1588 GAGCGCGCTTCGCCACCTCAAGAGT---GCGGTGGTGGCAACCTGACCGTCTCCACC 1644
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 767 ProAlaAlaHisSerIleLeuSerAlaProGlyAlaPheAspGlyLeuArgIleHisThr 786
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1645 AACTTCTTCATGAGGAGCTC-----AAAGCGCGACCTACCAAGGTGATTCTG 1695
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 787 AspGluIleGlnGluValLeuAlaArgPheGlnProGlyThrLeuThrValAlaVal 806
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1696 ATGACACACGTCGCTGCTGATATGCC-----1725
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 807 MetAspSerMetAspTrpPheAspProSerProGluGluGluLysGluGlyArgGly 826
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1726 GTGGCCAAAGAGTGGCGGAGTTCCTGCCCAAGCAGGTGGCGCGCGCATCGTCATC 1785
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 827 LysAlaArgGluGlnValArgArgLeuAsnArgAlaLeuLysValGlyGlyValLeu 846
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1786 TGGCGCTCGGCTCCTCAGCGCGCTTACGCGGAGTGTATCCAGAGCGCGGCTTCGAC 1845
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 847 LeuArgSerAlaGlyValGluProTrpTyrValArgValPheValGluGluGlyPheGly 866
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1846 GTGCGCTGCATC-----CGCGCGCCACTCAGGGCTACATGACGCGGTC 1890
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 867 AlaArgValGlyCysArgGluSerGlyArgGlyAspGlnGluCysIleAspArgVal 886
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1891 AACATGTACACTCTCTTACATGCGCGCGCGG 1923
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 887 AsnMetTyrAlaSerCysTrpIleLeuGluLys 897
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

## RESULT 4

```

US-10-084-846A-6
; Sequence 6, Application US/10084846A
; Publication No. US20040006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLERWEG, AGNES
; APPLICANT: TREPFER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084, 846A
; PRIOR FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: Patent In Ver. 3.2
; SEQ ID NO 6
; LENGTH: 19662
; TYPE: PRT
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:

```

```

; OTHER INFORMATION: Protein 1: amino acid sequence encoded by coding strand 2
; OTHER INFORMATION: Start codon: gga, Start position: nucleotide 1.
US-10-084-846A-6

```

```

Alignment Scores:
Pred. No.: 1,63e-20 Length: 19662
Score: 476.50 Matches: 262
Percent Similarity: 31.6% Conservative: 42
Best Local Similarity: 27.3% Mismatches: 272
Query Match: 13.2% Indels: 385
DB: 4 Gaps: 56

```

US-10-620-914-44 (1-1947) x US-10-084-846A-6 (1-19662)

```

QY 1945 AGTTGTCTCTTCTGGCGCCTTCGCGCGGCGCATGTAGAGAGAGCTGTACATGTTGAGCG 1886
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 12392 ThrCysAlaGlyArgArgProGlyArgAlaProAlaArgArgAlaCys-----Arg 12408
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1885 GGTCCATGTAGCCCTGAGTGGCGCGGATGACAGCGCACGTGGAAGC-----CCGCT 1832
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 12409 GlyGlySerProProArgSerArg---ArgCysArgProArgArgSerCysArgProGly 12427
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1831 TCTGATCAGCTCGCCTAGCGCGGCTGAGGAGGCGGAGCGGCACGATGACGATCGCGC 1775
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 12428 ThrArgSerArgArgArgGlySerHisGlyThrArgAlaArgArg-----Arg 12444
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1771 CCGCGCAACCTGCTGGCCAGGC-----ACTCGCCAGCTCGTGTGCCACGCCCATAT 1718
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 12445 ProAlaArgAlaGlyArgProAlaGlyCysArgArgProGlyArgThrProArgArgLeu 12464
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1717 CCAGCCAGTCCATGTCATCAGAAATCATTGTTAGTGTGGTGGCGCTTTGAGCTCCT 1658
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 12465 ProArgLysProGlySerProProArgSerPro-----12475
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1657 CCATGAAGAAGTTGGTGGAGACGCTCAGGTGTCCACCCACCCACTCTTGGAGGTGG---1601
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 12476 -----ArgArgGlnSerArgProProIlePheArgProProTrpAsn 12489
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1600 -----CGAAGC-----CCGCTCGCGCAGGTAGTGTCCACCCACCCACTCTTGGAGGTGG---12506
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 12490 ProValHisArgArgGluGlnPheTyrTrpProGlyProValTrp-----12506
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1564 CGCGCAGAACTTCCCGTGAGCGAGTGTAGTAGAAGTAGTCTGTCTGTCGCGACGTCG 1505
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 12507 -----ProArgTyrAlaSerAlaAlaAla 12514
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1504 AGTTCTCGGCACGCGCTCCA-----TGGTGGCGCGCATGT-----1469
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 12515 SerSerGlnProArgProIleAlaProSerTyrProThrProCysGlyAlaArg 12534
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1468 -----AGTTCTCAATGG-----1457
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 12535 GlnIleArgSerPheGlnGlyAsnSerArgTrpHisArgMetSerProProAsnAsnAsp 12554
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1456 -----GGATGCCGT-----1448
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 12555 ArgThrGlyValPhePheArgPheSerLeuGlyPheArgAlaGlyTrpArgGlyProPhe 12574
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1447 ---CCGCTTGATCAGCGCTACTGCTTGGCCGCGCAGCGCCCGACACACGACGACGG 1391
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 12575 ThrProProSerProAlaAspMetSerTyrGlyValSerAlaPheArgThrHisGlyArg 12594
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1390 CCTTGTGTAAGAGACCCAGC-----1370
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 12595 ProProSerGlyValProGlyValArgSerGlyAspArgArgCysArgTrpTrpAlaAla 12614
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1369 TCACGAATTTGACGAACAGCACACGAGCGGCTTGGCCGCTTCTTCCAGAGTGGATGA 1310
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 12615 SerArgArgCysTrpProAlaArgProGlyArgGlyAlaSerProAlaGlyGlyGlyPro 12634
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1309 GCATGTTGTCCTCCACAGAGCGGCTGCTCTCCATTG-----1271
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 12635 ProProSerSerProSerArgAlaProAlaProAlaGlyGlyGluArgAlaGlyPro 12654
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Qy 1270 TGGCGCGCTTGGCAGGCGCTTGACGGTCTTCCAGTCCAGCACCCAGCCAGGCACT 1211
Db 12655 TrpGlyArgValAlaSerProAlaArgSerAlaArgAlaProAlaArgAlaProGlyAsn 12674
Qy 1210 GCAGCA-----CCC 1202
Db 12675 AlaAlaProProAlaThrAspProArgThrProArgProProArgProArgArgPro 12694
Qy 1201 AGCAGAGTTGCCATGCCGC-----CCTGGTAGTACAGCGCGCTGCTGGA 1157
Db 12695 AlaArgArgCysProAlaArgProAlaAlaArgProProGlySerAlaGlyArgAlaGly 12714
Qy 1156 AGTACCAGAGCGCT----- 1142
Db 12715 SerProGlyValProProArgProProGlyArgAlaArgCysArgArgGlyAsp 12734
Qy 1141 -----TGG 1139
Db 12735 ArgAlaArgProArgProSerProProValAlaCysArgArgArgSerArgArgArg 12754
Qy 1138 ACCAGAGTTGGCTGCTTT-----GCCAGAGGAAG 1106
Db 12755 AlaArgProCysGlyAlaAlaGlyProArgArgProAlaArgSerThrSerArg 12774
Qy 1105 GCGCCAGCTTCTCT-----CGTACA-----GCTCTCAATGCGCGGTGCA 1064
Db 12775 ArgProAlaAlaGlyProArgThrArgArgAlaArgAlaProGluArgAlaValHis 12794
Qy 1063 CGCCCTCCGCCAAGCTGCC-----ACAGTCTCTCAA 1031
Db 12795 ArgAspArgGThrValAlaArgArgSerAsnArgAsnArgSerArgValArgProPro 12814
Qy 1030 ACTCAGAGTGTGAATGGCCACTTTCAGCTCCAGAGCGCCGACTCGCGG---GGT 974
Db 12815 ArgProAlaAlaGlyArgProGlyArgGlyArgProValArgAlaGlyAlaArgProGly 12834
Qy 973 TGCAGTCCACCGACACCACTGGCCGG-----CCCTTCGCCAGCAGAGTTCAGGGAT 920
Db 12835 ProAlaAlaThrAlaProProAspArgArgGlyProAlaAlaArgProSerGlyProArgHis 12854
Qy 919 ---TGCAGCGCGCTAGTCAGGG---TCAGCACCGTGTCTTGGGTTGATCTCCATCA 866
Db 12855 GlyArgAlaArgGlySerGlyArgArgArgProArgProSerGlyArgProAlaAla 12874
Qy 865 CTCCATATCCGGCTCGGGCTCTCCACGACTGCTGTACAGGAAGTGGCGGGAACA 806
Db 12875 ProProArgPro---ArgGlyArgPro-----GlySer 12884
Qy 805 TGGCGCGCGCTCCAGCGCTCTCTGTGACAGGCGGTGGCCACGCTGGCGAGCGCCNA 746
Db 12885 AlaAlaAspProProSerAlaPro-----GlyGlySer 12895
Qy 745 TCCACAGTGTAGTGGGGCGCGCAGCCAGCGTAGGGGATCGAAACCTGGGTGTTCT 686
Db 12896 SerSerCysProGlyGlyArgGlyAlaLeuArgProArgProArgProProGly---Thr 12914
Qy 685 GTCCACACGGGCTCCAGCTTCTGCTCCAGTAGG---CGCGGCGCTCGGGG--- 635
Db 12915 ProProAlaArgProGlyArgSerProArgArgGlyArgGlyCysProGlyThrGly 12934
Qy 634 -----CGATGTCATGTTGT-----CGATGT-----CGAAGA 608
Db 12935 GlyCysArgArgProAlaCysCysAlaThrAlaAlaGlyArgCysProArgArgArg 12954
Qy 607 TCGATC-----GCCAGAGAAACGGCGCACACAGG----- 578
Db 12955 AlaCysProAlaAlaCysGlySerArgArgArgGlyArgAlaArgProLeuAlaProArg 12974
Qy 577 -----GCATCTGGCGAGGGCGGAGGTCTGCTACTTGC----- 548
Db 12975 ArgTrpAsnAlaThrAlaGlyGlyAlaAlaArgGlyArgAlaAlaGlyLeuArgGlySer 12994

```

---

```

Qy 547 -----CGTCACTAGTAGCTCGGCCAACGCCACCCAGGC 515
Db 12995 ArgSerAlaGlyAlaProThrSerProArgSerArgGlySerGlyArgArgProArg 13014
Qy 514 CGTCTTGGGACAGTACGAGCAAGCCTGCTGATGAGCTTGT-----GGAAACGGTGAA 461
Db 13015 ArgArgGlyArgSerGlyProArgAlaGly-----CysAlaAlaAlaSerValArg 13031
Qy 460 TCATCGTGGCGAGTACGAGGAGGTGATGAGCGTCGCGG-----TGCCCTCAG 413
Db 13032 ArgThrArgProProArgArgValArgProAlaSerArgGlySerSerAlaArgProArg 13051
Qy 412 GGGCGCAAAATTGCCAAGCGTCGCGCT-----CCACGACCTGGACATCTCT----- 368
Db 13052 GlyAlaGlyThrAlaArgArgThrProGlyArgProArgProArgSerArgArg 13071
Qy 367 -----TCAGGCCCTTGG 356
Db 13072 ArgAlaAlaProGlyCysAlaGlySerProProSerThrArgArgArgSerProTrp 13091
Qy 355 CCT-----TCGCCCTCT-----TCTTGCCA----- 335
Db 13092 ProAlaGlyProArgAlaProSerCysArgAlaAlaCysAlaAlaArgTrpAspSerAla 13111
Qy 334 -----CCTCGCACAGCG-----AGTGGCACAGGTCCGACA 305
Db 13112 ArgArgArgProArgAlaAlaGlyArgAlaGlyGlyProCysAlaGlyCysGlyArgPro 13131
Qy 304 CGT----- 302
Db 13132 ArgArgArgProAlaValArgProAlaProArgArgSerSerGlyArgArgProSer 13151
Qy 301 -----AGATGGACTTGAATTCAGCTCGCCAGGTGATGTAAT 269
Db 13152 ArgProArgArgGlyHisArgArgSerArgTrpThr-----SerSerArgThrArgTrpPro 13170
Qy 268 CAGCCATCATATCGACATTTCTCCCGAGTGC----- 239
Db 13171 ArgProPro---ArgHisCysProArgAlaAlaAlaArgArgProArgArgAlaAspPro 13189
Qy 238 -----CACCCAGGTCAACCCAGAGTAGAGTTTCGAGC-----GCTCGGCCAGGC 194
Db 13190 SerProSerArgHisAlaProArgProArgAlaAlaAlaAlaThrTyArgSerAla 13209
Qy 193 GGGCAGCAAGAGCAGCGCGCTGGGCGCTAGAACCTCTCCAGCGAGCAGCGTGTATCAT 134
Db 13210 GlyProArgSerArgSerAlaProArgArg-----ArgArgArgSerHis 13226
Qy 133 CGCCCTTCT-----TGTCGCGAACCACA 110
Db 13227 ArgProSerGlyArgArgArgAlaArgArgProValThrAlaThrAlaArgAla 13246
Qy 109 TATGCGCGAGAACCG-----TCAGGT-----CATCTTTCATGCTGCTGAGCTTGA 65
Db 13247 ArgAlaAlaSerArgArgSerTrpSerGlyArgArgArgProArgCys----- 13262
Qy 64 GCTTCTCCAGGAGAGTCTTCTTGTGTACTCTCGAGCGCGCGCTCAGCACCCGAGCC 5
Db 13263 -----GlyArgProAlaAlaSerAlaGlyAlaArgThrArgSerProProThr 13279
Qy 4 CCA 2
Db 13280 Pro 13280

```

RESULT 5  
US-10-084-846A-5  
; Sequence 5, Application US/10084846A  
; Publication No. US20040006026A1  
; GENERAL INFORMATION:  
; APPLICANT: WEITNAUER, GABRIELE  
; APPLICANT: MUHLNBERG, AGNES  
; APPLICANT: TREFFER, AXEL  
; APPLICANT: BECHTHOLD, ANDREAS



; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES  
 ; FILE REFERENCE: 1974-005  
 ; CURRENT APPLICATION NUMBER: US/10/084,846A  
 ; CURRENT FILING DATE: 2003-02-25  
 ; PRIOR APPLICATION NUMBER: PCT/EP01/09815  
 ; PRIOR FILING DATE: 2001-08-24  
 ; PRIOR APPLICATION NUMBER: DE 101 09 166.4  
 ; PRIOR FILING DATE: 2001-02-25  
 ; NUMBER OF SEQ ID NOS: 120  
 ; SOFTWARE: PatentIn Ver. 3.2  
 ; SEQ ID NO 5  
 ; LENGTH: 19723  
 ; TYPE: PRT  
 ; ORGANISM: Streptomyces viridochromogenes  
 ; FEATURE:  
 ; OTHER INFORMATION: Protein 3: amino acid sequence encoded by coding strand 1.  
 ; OTHER INFORMATION: Start codon: atc, Start position: nucleotide 3.  
 US-10-084-846A-5

Alignment Scores:  
 Pred. No.: 1.95e-19 Length: 19723  
 Score: 459.00 Matches: 241  
 Percent Similarity: 30.8% Conservative: 48  
 Best Local Similarity: 25.7% Mismatches: 271  
 Query Match: 12.7% Indels: 378  
 DB: 4 Gaps: 50

US-10-620-914-44 (1-1947) x US-10-084-846A-5 (1-19723)  
 QY 32 GCTACACCAAGAGAACTTCTCCCTGGAGAGCTCAAGCTCAGCAGCATGAGCATGACC 91  
 DB 6473 AlaArgProGlyGlyValAlaArgTrpArgTrpSer----- 6484  
 QY 92 TGACCGTTCTCGCCATATGTGGTTTC-----GCAGCAAGAGGGCGCATGATC 139  
 DB 6485 -----ValCysProSerThrArgProProSerGlySerArgAlaMet--- 6498  
 QY 140 ACGTGTCTGCTGGAGAGCTTCTACGGGCCCCAGG-----CGCTGCTCTTGTCTG 190  
 DB 6499 -----ThrArgTrpSerArgProSerArgProArgArgProProSerProValThr 6516  
 QY 191 CCCGCTGGCCGAGCGCTCGAACCTCATCTGGTTGACCTGGTGGTGGCACTGGGAGA 250  
 DB 6517 ProAlaArgSerArgArgSerAlaAlaGlyProThrArgGlyTrpSerTyrGlyGly 6536  
 QY 251 ATGTCGATATGATGCTGATTCATCATCGACCTGGCGAAGTTCAGTCCATCTACGTGTCG 310  
 DB 6537 Ser-----ProProThrTrpProSerThrAla---ArgThrTrpThr 6549  
 QY 311 ACCTGTGCCACTCGCTGTGGAGTGGCCA----- 340  
 DB 6550 ProCysSerThrGlyCysAlaArgProProThrSerValProAlaHisArgThrArgArg 6569  
 QY 341 -----AGAGAGCGGAGCCAGCGCTGGAGA-----ATGTCAGGTCTGGAGG 388  
 DB 6570 SerThrArgArgArgSerArgThrArgGlyProArgProProThrAlaHisArgThrGlyArg 6589  
 QY 389 CCAGCGCTTGCATTTG-----CGCCCTGAGGCGACCGCGACGCTCATCACCTTCT 442  
 DB 6590 ArgThrCysAlaThrArgCysGlySerProProArgSerProArgArgProArgThrAla 6609  
 QY 443 CTTACTCGCTCAGCATGATTCACCGTTCCACAGCTCATCGACCGCTTGTCTGCTGACC 502  
 DB 6610 ThrGlyArgSerSerCysProArgThrProTrpSerArgThrArgSerThrArgArg 6629  
 QY 503 TGTCCCAAGACGGCTCG----- 520  
 DB 6630 TrpProAlaAlaTrpArgThrCysSerSerAspArgArgTyrGlyAlaThrSerPro 6649  
 QY 521 -----TGGCGGTGGCGACTTCTACGTGAGCG----- 547  
 DB 6650 ArgProGlyProSerAlaProProTrpAlaProProThrAlaThrAlaAlaTrpThrGly 6669

QY 548 -----GCAAGTACGACCTGCCCCCTGCCAGCATGCCCTGGTGGCCGCGTT 592  
 DB 6670 ArgArgSerSerArgThrGlyThrSerLysSerCysArgProThrProGlySerThrAla 6689  
 QY 593 TCTTCTGCG-----GATCGATCTTCGACATCGACACATTGACATCG 634  
 DB 6690 ArgCysGlyValProSerProGlyArgGluArgProSerAlaAlaThrThrSerThr--- 6708  
 QY 635 GCCCGAGCGCGCGCTTACCTGCGAGCAGAAGTGGAGCGGTGGGAGGAGCACACCC 694  
 DB 6709 -----ProThr-----ArgCysTrpAlaArgProAla 6717  
 QY 695 AGGTTGATCCCTACGTGCGTGGTGGCGCCCTACTACTACGTGTGGATTGCCGCC 754  
 DB 6718 AlaSerArgAlaAlaThrCys---GlyCysGlyThrAlaArgTrpThrThrAlaAla 6736  
 QY 755 TGCCAGCGTTGGCCAGCGCTCGCAGGAGCGGTGGAGCGCGCCGCGCATGTTCCGCG 814  
 DB 6737 ArgThrArgAlaAlaThrProSerThr-----AlaTrpArgSer-----SerArg 6751  
 QY 815 CCACCTTCTCTACACGCGAGTGGGGAGGACCCGCGCGATATGAGGTGAGGAGA 874  
 DB 6752 ProPro---CysTrpArgSerArgSer-----TrpArg 6761  
 QY 875 TCAACCCCAAGGACACCGTGTGCTGACCTGACGCGCGGTGCAATGCCCTGAACCTGCG 934  
 DB 6762 ProAlaProArgAlaLysSer-----AlaAlaProSerArgThrArgProThr 6777  
 QY 935 TGGTGCAGG-----GGCGCGCGCAGGTGTGTGCGTGG 967  
 DB 6778 ArgCysArgArgValSerGlyArgSerArgSerSerAlaArgAlaArgTrpCysGlyTrp 6797  
 QY 968 -----ACTGCAACCCCG----- 979  
 DB 6798 ProProGlyArgSerArgThrProProThrArgThrProProGlySerSerMetProArg 6817  
 QY 980 -----CGCAGTGGCGCG----- 991  
 DB 6818 ProGlyArgProArgArgThrSerProValTrpArgArgGlyArgCysTrpThrArgAla 6837  
 QY 992 -----TTCTGGAGCTGAGAGGTGGCGCTTTCAGGCG 1024  
 DB 6838 SerThrGlySerAsnArgProThrProAlaTrpSerProGlyGlyTrpPro-----Arg 6855  
 QY 1025 TGGAGTTTGGAGCGTGTGCGAGCTGTTCGCGAGGCGG---TGCACCGCGCATTGAGG 1081  
 DB 6856 TrpAlaTrpProArgProValSerIleGlyAlaSerSerAspCysProProValSerVal 6875  
 QY 1082 AGCTGTACGAGAAGAGCTGCGCGCTTCTGTGCGAAACAGCCACAACTTCTGTGCTCA 1141  
 DB 6876 TyrCysThrLeuArg-----CysSerArgProThrProArgProGlyPro 6890  
 QY 1142 AGCGCTCTGTTACTCCAGCAGCGCTGTACTACAGG-----GGGCGATGGCGAAGC 1195  
 DB 6891 ArgCys-----TrpThrProCysArgSerArgArgProProSerTrpAlaSer 6906  
 QY 1196 TGTGCTGGGTGCTGCTGCTGCGTGTGGAGTGGGAGTGGGCAAGACCTCAAGC--- 1252  
 DB 6907 ArgSer-----SerAlaTrpSer---CysThrSerThrArgSerProSerThrAla 6922  
 QY 1253 -----GCCTCGCCCAACG----- 1264  
 DB 6923 ArgHisArgArgArgArgSerArgSerArgSerIleProAlaSerProThrProCys 6942  
 QY 1265 -----CGCCCAATGAGGAGGAGCGCGCTGTGGGACA 1300  
 DB 6943 ThrProTrpSerArgThrGlyArgAspAlaArgTrpArgAlaCysAlaAlaThr 6962  
 QY 1301 GCAACATGCTCATCTCGTGAAGAAACCGGCGCCCAAGCGCG----- 1342  
 DB 6963 ArgTrpSerSerSerArgProArgArgThrProThrSerArgAlaAlaThrArgThrArg 6982  
 QY 1343 ---TGGTGT-----GGCTGTTCG----- 1357

Sequence 5, Application US/10084846A  
Publication No. US20040006026A1  
GENERAL INFORMATION:  
APPLICANT: WEITNAUER, GABRIELE  
APPLICANT: MUHLENWEG, AGNES  
APPLICANT: TREFZER, AXEL  
APPLICANT: BECHTHOLD, ANDREAS  
TITLE OF INVENTION: AVILAMYCIN DERIVATIVES  
FILE REFERENCE: 1974-005  
CURRENT APPLICATION NUMBER: US/10/084,846A  
CURRENT FILING DATE: 2003-02-25  
PRIOR APPLICATION NUMBER: PCT/EP01/09815  
PRIOR FILING DATE: 2001-08-24  
PRIOR APPLICATION NUMBER: DE 101 09 166.4  
PRIOR FILING DATE: 2001-02-25  
NUMBER OF SEQ ID NOS: 120  
SOFTWARE: PatentIn Ver. 3.2  
SEQ ID NO 5  
LENGTH: 19723  
TYPE: PRT  
ORGANISM: Streptomyces viridochromogenes  
FEATURE:  
OTHER INFORMATION: Protein 3: amino acid sequence encoded by coding strand 1.  
OTHER INFORMATION: Start codon: atc, Start position: nucleotide 3.  
US-10-084-846A-5

Alignment Scores: 2.41e-19 Length: 19723  
Pred. NO.: 457.50 Matches: 242  
Score: 33.1% Conservative: 46  
Percent Similarity: 27.8% Mismatches: 256  
Best Local Similarity: 12.6% Indels: 325  
Query Match: 4 Gaps: 55  
DB:

US-10-620-914-44 (1-1947) x US-10-084-846A-5 (1-19723)

QY 1945 AGTGTCTCTTCGCGCCCTTCGCGCGGCGCATGTAGA-----1907  
DB 6222 SerThrSerGlyTTPArgPro-----GlyProCysArgGlyProSerValArgThrPro 6239  
QY 1906 -----AGAGCTGTACATGTTGACGCGTCCATGTAGC-----CCT 1871  
DB 6240 AlaGlyArgLeuCys-----ProValSerAlaThrAlaAlaProThrPro 6254  
QY 1870 GAGTGGCGCGCGATGACGACGCTCGAAGCGCCTTCGTGATCAGCTCGCGTAGG 1811  
DB 6255 MetTrp-----SerSerAsnArgArgProProArgProProAlaProArgArg 6270  
QY 1810 GCGGCTGAGGAGGCGGAGCGCCAGATGACGATGCCCGCGCGCAACCTCTCTGGCCA 1751  
DB 6271 Ser-----ArgPro-----GluProCysSerArg 6278  
QY 1750 GGCATCGGCCAGCTCGTTGGCCAGGCGCATATCCAGCAGTCCACGTGTCATCAGAA 1691  
DB 6279 CysProArgProProArg---ArgArgSerValThrAlaArgProArgSerProSerGly 6297  
QY 1690 TCACCTTGGTGTAGTGGCGCTTTGAGCTCTCCATGAAGAAGTTGGTGGAGACGGTCA 1631  
DB 6298 Ser-----ArgArgAlaLeu-----ThrTrpProArgSer 6307  
QY 1630 GGTGTGTCCA-----CCAGCCCACTCTTAGGG-----TGGCGAAGG 1595  
DB 6308 GlyThrProTrpProIleGlySerProArgTrpSerThrGlyArgProSerTrpArgPro 6327  
QY 1594 -----CCGCTCGCGCAGGTAGTGGGC-----1571  
DB 6328 AlaGlyThrAsnTrpProProGlyCysAlaArgTrpProLeuArgAsnProArgProVal 6347  
QY 1570 AGTGTGCGCGAGCACTTGGCGGTGAGGAGTGTAGTAGAAGTAGTAGTCTGCTCGCA 1511  
DB 6348 SerSerProGlyProArgCysArg-----MetProAlaValProSerGlyCysSer 6364  
QY 1510 CGTGCAGGTCTTCGCCACCGCGCTCCATGGTGGCGCGATGT-----1469

6983 ThrTrpCysArgSerArgValCysArgArgSerCysValProGlyCysSerThrArg 7002  
QY 1358 -----TCAAGTTCGTGAGCTGTGTCTTCA---ACAAGGCGTGTCTGT-----1399  
DB 7003 CysAlaSerArgSerArgArgCysAspSerProProArgProCysThrPheAlaAla 7022  
QY 1400 -----GGTTCGGCGCGCGGTGCGCGCAAGC 1426  
DB 7023 ArgTrpTrpSerArgGlySerThrArgArgTrpTrpSerAlaAlaGlyTTPArgSerAla 7042  
QY 1427 AGTACGCGCTGATCAAGCGGACGGCATCC-----CCATTGAGACTACATCG 1474  
DB 7043 SerAlaGlyThrCysArgProTrpSerSerGlySerCysProProSerAlaThrSerSer 7062  
QY 1475 CGCGCACCA-----TGGACGCGTGGCGGAGAACTCGCACGTGCGCAAGCAGAACT 1525  
DB 7063 IleThrProAsnAlaSerArgAsnThrArgArgArgThrAlaMetArgProGluArgThr 7082  
QY 1526 ACTTCTACTACACTGCTCAACGGCAAGTTCTTCGCGCACAACTGCGCCCACT-----1579  
DB 7083 AspSerValThrHisValSerArgGlyArgGlyGluArgAlaAlaProAlaGlyPro 7102  
QY 1580 -----ACCTGCGGAGCGCGCTTCGCCA 1603  
DB 7103 GluThrGlyLysProGluLysHisAlaSerThrIleValValGlyArgArgHisSerMet 7122  
QY 1604 CCCTCA-----1609  
DB 7123 ProSerGlyValProLeuLysArgProAspLeuThrAlaLeuThrAlaArgCysGlyIle 7142  
QY 1610 ---AGAGTGGCGTGTGACCAACTGACCGTCTCCACCACTTCTTCATGAGAGAGCTCA 1666  
DB 7143 ArgArgSerTyTrpTrp-----AlaGlyLeuAlaArgGlyCys 7155  
QY 1667 AAGCGCGACCTACACCAAGGTGATCTGTATGACCACTGCG-----1708  
DB 7156 GlyArgLeuArgIleProArgProAsnAsnTrpSerArgTrpProIleGluLeuPheSer 7175  
QY 1709 ---ACTGGCTGATA-----1720  
DB 7176 ProMetTyTrpIleProSerArgArgProGluTyArgArgSerGlyLeuSerSerGly 7195  
QY 1721 -----TGCCCGTGGCCAAAGAGTGGCGCGAGTGGCGCTGCTGGCCAAAGC 1759  
DB 7196 ArgProArgArgSerArgLeuProTrpGluThrAlaArgSerSerAlaTrpThrSer 7215  
QY 1760 AGTTGCGCGCGCGCATCGTCATCTGGCGTCCGCTCCCTCAGCGCGCCCTACGCG 1819  
DB 7216 ProThrSerSerArgThrProArgAlaGlyTrpProThrProCysSerCysProValThr 7235  
QY 1820 AGCTGATCCAGAAGCGCGCTTCGACGTGC-----1849  
DB 7236 Ser---SerAlaThrThrSerSerCysThrTrpAlaThrThrSerTrpThrAla 7254  
QY 1850 -----GCTGCATCCGCGCGCCACTCAG-----1873  
DB 7255 SerSerThrSerTrpArgThrSerAlaAlaThrSerProProArgArgSerCysSerPro 7274  
QY 1873 -----1873  
DB 7275 GlySerProThrArgAlaGlySerAlaTrpProAsnLeuThrArgThrAlaGlyProArg 7294  
QY 1874 -----GCTACATGACCGCGCTCAACATGTACA-----1900  
DB 7295 SerTrpArgSerProArgThrProGlyAlaThrTrpArgSerSerAlaCysThrProSer 7314  
QY 1901 ---GCTCCTTCTACA-----TGGCGCGCGGAAAGCGCGCCAGAGAGCAACT 1945  
DB 7315 AlaArgProCysThrArgProTrpProAlaSerGlyArgProGlyGlyThrSer 7332  
RESULT 6  
US-10-084-846A-5

6365 ArgValThrGlyProSerGlyProGlyTrpAlaAlaAsnCysTrpLysProSerArgSer 6384  
||| :||| ||| ||| ||| |||  
1468 ---AGTTCATGG-----GGATCGCGTCCGCT 1442  
||| :||| ||| ||| ||| |||  
6385 SerProArgSerThrAsnTrpSerArgCysSerArgArgSerGlySerArgPro--- 6403  
||| :||| ||| ||| ||| |||  
1441 TGATCAGCGGTACTGCTGCGCGCAGCGCCGCCGACACACAGCA---CGGCTTGT 1385  
||| :||| ||| ||| ||| |||  
6404 ---GlyArgCysTrpArgAlaThrTrpProArgSerThrAlaProArgProSer 6421  
||| :||| ||| ||| ||| |||  
1384 TGAAGACACCA---GCTCAGCACTTGACGAACAGCCACACCA 1343  
||| :||| ||| ||| ||| |||  
6422 SerArgCysSerGlySerProArgCysGlyGlyHisGlyAlaSerSerProPro 6441  
||| :||| ||| ||| ||| |||  
1342 GCGCTTGGGCGCTTCTTCAGAGT----- 1316  
||| :||| ||| ||| ||| |||  
6442 SerSerAlaThrArgSerValArgSerProProSerProProGlyArgProProThr 6461  
||| :||| ||| ||| ||| |||  
1315 ---GGATGAGCA---TGTTGCTGTCCACAGAGCGGCTGCTCTCTCCATG 1271  
||| :||| ||| ||| ||| |||  
6462 GlyProGlySerAlaValGlyArgCysCys---AlaArgProArgGly 6477  
||| :||| ||| ||| ||| |||  
1270 TGGGCGCTTCGCGGCGCTTGACGGTCTGCCAGTCCAGCACCA--- 1223  
||| :||| ||| ||| ||| |||  
6478 ValAlaArgTrpArgTrpSer---ValCysProSerThrArgProProSerGlySer 6495  
||| :||| ||| ||| ||| |||  
1222 ---CGGCGGCGCTGCGAGCACCCAGCACAGCT 1193  
||| :||| ||| ||| ||| |||  
6496 ArgAlaMetThrArgTrpSerArgProArgProArgArgProProSerProVal 6515  
||| :||| ||| ||| ||| |||  
1192 TGCCCATCCCGCTGTAGTACAGCGTGTGGAAGTACAGAGCGCTTGACACAGA 1133  
||| :||| ||| ||| ||| |||  
6516 ThrProAlaArgSerArgSerAlaAlaGlyProThrArgGly---TrpSerTyr 6534  
||| :||| ||| ||| ||| |||  
1132 AGTTGTGC---TGCTTTCGACAGGAGG--- 1106  
||| :||| ||| ||| ||| |||  
6535 GlyGlySerProProThrTrpProSerThrAlaArgTrpTrpProCysSerThrGly 6554  
||| :||| ||| ||| ||| |||  
1105 ---CGCGCCAGCT---TCTTCTCGTACAGCT 1082  
||| :||| ||| ||| ||| |||  
6555 CysAlaArgProProThrSerValProAlaHisArgThrArgSerThrArgArg 6574  
||| :||| ||| ||| ||| |||  
1081 CCTCAATCGCGGTGACAGCCCT---CGCCGACAGCTGCCACA 1040  
||| :||| ||| ||| ||| |||  
6575 SerArgThrArgGlyProArgProProThrAlaArgThrGlyArgThrCysAlaThr 6594  
||| :||| ||| ||| ||| |||  
1039 CGT---CCTCAACTCCA---GCTGTGATGCCACCTTCT 1004  
||| :||| ||| ||| ||| |||  
6595 ArgCysGlySerProProArgSerProArgProArgProAlaThrGlyArgSerSer 6614  
||| :||| ||| ||| ||| |||  
1003 TCAGCTCCAGAACGCGCGACTGCG--- 980  
||| :||| ||| ||| ||| |||  
6615 SerCysProArgThrProTrpSerArgThrArgSerThrArgTrpProAlaAla 6634  
||| :||| ||| ||| ||| |||  
979 ---CGGGTTCAGT---CCACCG 962  
||| :||| ||| ||| ||| |||  
6635 TrpArgThrCysSerSerArgArgTyGlyAlaThrSerProArgProGlyProSer 6654  
||| :||| ||| ||| ||| |||  
961 ACACACTGCGCGCGCCCT---GCACCAGGTTGAGGCAATGACGCGCGC 908  
||| :||| ||| ||| ||| |||  
6655 AlaProProTrpAlaProProThrAlaThrAlaAlaTrpThrGlyArgSerArg 6674  
||| :||| ||| ||| ||| |||  
907 TAGTCAGGTCAGCACCGTGTCTTGGGTTGATCTCCATCACT--- 863  
||| :||| ||| ||| ||| |||  
6675 ThrGlyThrSerLysSerCys---ArgProThrProGlySerThrAlaArg 6690  
||| :||| ||| ||| ||| |||  
862 ---CATATCCGCTCGGGTCTTCCACAGCTCGGTACAGGAAGTGGCG 812  
||| :||| ||| ||| ||| |||  
6691 CysGlyValProSerProGlyArgGluArgProSerAlaAlaThrThrSerThrProThr 6710  
||| :||| ||| ||| ||| |||  
811 GGAACATGGCG---CGGCTCCACGCGCTCTCTGTCAGGGCGCTGGC---CAACGC 761  
||| :||| ||| ||| ||| |||

6711 ArgCysTrpAlaArgProAlaAlaSerArgAlaAlaThrCysGlyCysGlyThrAlaArg 6730  
||| :||| ||| ||| ||| |||  
760 TGGGCGCGCGCAATCCACACAGTAGTAGGGCGCGCAGCC---ACGCGACGT 710  
||| :||| ||| ||| ||| |||  
6731 TrpThr---ThrThrAlaAlaArgThrArgAlaAlaThrProSerThrAlaTrp 6747  
||| :||| ||| ||| ||| |||  
709 AGGGATCGAACCTGGGTGT---TCTGTCCCAACACGCGT 671  
||| :||| ||| ||| ||| |||  
6748 ArgSerSerArgProProCysTrpArgSerArgSerTrpArgProAlaPro---ArgAla 6766  
||| :||| ||| ||| ||| |||  
670 CCAGCTTCTGCTCCAGGTAGCGCGCTCGGGCCGATGTCATGTGTGCGATCGA 611  
||| :||| ||| ||| ||| |||  
6767 LysSerAlaAlaPro---SerArgThrArgProThrArgCysArg 6780  
||| :||| ||| ||| ||| |||  
610 AGATCGATCGCAGAGAAACGCGCGACAGGCGATCTGGCGCAGG---GCAGGT 557  
||| :||| ||| ||| ||| |||  
6781 ArgValSerGlyArgSerArgSerSerAlaArgAlaArgTrpCysGlyTrpProGly 6800  
||| :||| ||| ||| ||| |||  
556 CGTACTTCCGCTCACGTAGAAAGTGGCAACGCCACCGCCGCTTTGGGACAGTAGC 497  
||| :||| ||| ||| ||| |||  
6801 ArgSer---ArgThrProProThrArgThrProProGlySer 6813  
||| :||| ||| ||| ||| |||  
496 AGC---AAGCTGTGATGAGTGTGCGAACGGTGAATCATCTGAGCAGTAGG 443  
||| :||| ||| ||| ||| |||  
6814 SerMetProArgProGlyArgProArg---ArgThrSerProValTrpArg 6829  
||| :||| ||| ||| ||| |||  
442 AGAAGGTGATGAGCGTCCGCTGCGCTCAGGGGCGCAAAATTGGCAAGCTCGGCTCCA 383  
||| :||| ||| ||| ||| |||  
6830 Arg---GlyArgCysTrpThrArgAlaSerThrGlySerAsnArgPro--- 6844  
||| :||| ||| ||| ||| |||  
382 CGACCTGGACATCTTCCAGCCCTTGGCTTCTTCTTGGCCACCTCGCACAGC 323  
||| :||| ||| ||| ||| |||  
6845 ---ThrProAlaTrpSerProGlyGlyTrpPro--- 6854  
||| :||| ||| ||| ||| |||  
322 AGTGGCAGCTGCACACAGTAGAGTGAACCTTGGCCAGCT--- 278  
||| :||| ||| ||| ||| |||  
6855 ---ArgTrpAlaTrpProArgProValSerIleGlyAlaSer 6867  
||| :||| ||| ||| ||| |||  
277 ---CGATGTAATCAGCCATCATATCGA 254  
||| :||| ||| ||| ||| |||  
6868 SerAspCysProProValSerValTyCysThrLeuArgCysSerArgProThrProArg 6887  
||| :||| ||| ||| ||| |||  
253 CATTTCTCCAGTCCACACCCAGGTCAACCCAGATGAGGTTCGAGCGCTCGGCCA--- 197  
||| :||| ||| ||| ||| |||  
6888 ProGlyProArgCysTrpThrProCysArgSerArg---ArgProPro 6902  
||| :||| ||| ||| ||| |||  
196 ---GGC 194  
||| :||| ||| ||| ||| |||  
6903 SerTrpAlaSerArgSerSerAlaTrpSerCysThrSerThrArgSerProSerThrAla 6922  
||| :||| ||| ||| ||| |||  
193 GGGCAGCAAGGCGAGCGCGCTGGGCGCCGTAGAGCTCTCCAGCGGAGCAG--- 143  
||| :||| ||| ||| ||| |||  
6923 ArgHisArgArgArgArgSerArgSerArgSerIleProAlaSerProThrProCys 6942  
||| :||| ||| ||| ||| |||  
142 ---CGTGATCATCGCCCTTCTTCTGCTCCGCAACACCATATGCG 104  
||| :||| ||| ||| ||| |||  
6943 ThrProTrpSerArgThrGlyArgAspAlaArgTrpArgAlaCysAlaAlaCysAlaThr 6962  
||| :||| ||| ||| ||| |||  
103 GCAGAACGTCAGGTATCTTCTCATGCTGCTGAGCTTGAGCTTCTCCAGGAGAGTTCT 44  
||| :||| ||| ||| ||| |||  
6963 ArgTrpSerSerArgProArgTrpProThrProThrSerArgAlaAlaThrArgThrArg 6982  
||| :||| ||| ||| ||| |||  
43 TCTGTGTAGTCGACGCGCGCGCT 17  
||| :||| ||| ||| ||| |||  
6983 ThrTrpCysArgSerArgValCysArg 6991  
||| :||| ||| ||| ||| |||

## RESULT 7

US-10-084-846A-7  
; Sequence 7, Application US/10084846A  
; Publication No. US2004006026A1  
; GENERAL INFORMATION:  
; APPLICANT: WEITNAUER, GABRIELE  
; APPLICANT: MUHLENWEG, AGNES

```

; APPLICANT: TREFFER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084, 846A
; CURRENT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 7
; LENGTH: 19652
; TYPE: PRT
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: Protein 2: amino acid sequence encoded by coding strand 2.
; OTHER INFORMATION: Start codon: gat, Start position: nucleotide 2.
US-10-084-846A-7

Alignment Scores:
Pred. No.: 2,77e-19 Length: 19652
Score: 456.50 Matches: 232
Percent Similarity: 36.3% Conservative: 38
Best Local Similarity: 31.2% Mismatches: 247
Query Match: 12.6% Indels: 227
DB: 4 Gaps: 50

US-10-620-914-44 (1-1947) x US-10-084-846A-7 (1-19652)
QY 1945 AGTTGCTCTCTGGCGCCCTTCGCGCGGCATGTAGAGAGCTGTACATGTTGACGC 1886
DB 1918 AlaCysProAlaThrLeuProPheProPro-----1928
QY 1885 GGTCCATGTAGCCCTGAGTGG-----1865
DB 1929 GlyProAlaHisProLeuTriLeuAlaProLeuArgThrProAlaArgProAlaGlyLeu 1948
QY 1864 -----CGCGGCGGA 1856
DB 1949 LeuProProLeuAlaArgProArgProAlaProArgValProLeuProValArgProGly 1968
QY 1855 TGCAGCGCAGCTCGAAGCCCG---CCTTCTGATCAGCTCGG-----1817
DB 1969 -----ThrArgArgThrProArgProSerAlaArgSerArgArgProGlnArgSer 1986
QY 1816 CQTAGGCGGGGTAGGGAGGCGGAGCGGCAGATGACGATGCGCGCGCGCAACTGCT 1757
DB 1987 ArgAlaAlaGlyAlaProSerArgThrAlaPro---AspAlaArgThrProProCysAla 2005
QY 1756 TGGCCAGGCACTCGGCAGCTGTTGGCCAGCGGCATATCCAGCCAGTCCAGTGTCCA 1697
DB 2006 SerProSerCysArgArgAlaArg---ProArgGlyArgPro-----ProArgProPro 2022
QY 1696 TCAGATACCTTGGTGTAGGTGCGCGCTTTCAGCTCTCCATGAGAAAGTTGGTGAGA 1637
DB 2023 CysTrpAlaProProThrArgProCysAlaGlyAlaProGlyValArgGly-----Arg 2040
QY 1636 CGGTCAAGTGT-----1625
DB 2041 ProAlaGlyCysGlyAlaAspArgProValArgProProCysGlyThrProTrpSerPro 2060
QY 1624 CCACACGCACTCTTGAGGGTGGCGNAGCGCCCTCGCGCAGTAGTGGGCGAGTGT 1565
DB 2061 ProArgArgArgAlaArgArgTrpProProProAlaArgAlaGly-----GlySerAla 2078
QY 1564 ---CGCGCAGAACTTGGCGGTGAGGCAGT-----TGTAGTAGAAGTAGTCTGCTTGC 1514
DB 2079 ProGlyCysGlyThrProArgArgGlySerProTrpAlaTrpArgSerGlyThrAlaArg 2098
QY 1513 GCACGTGCGAGTTCTCGCCA-----CGCGTCCATGTGGCGCGATGT-----AGT 1466

```

```

2099 GlyArg---SerAlaProProThrArgArgProTrpSerGlyGlySerGlyAspSer 2117
1465 TCTCAATGGGATGCGCTCGCTTGTATCAGCGCGTACTGCTTGGCCGCGCAGCGCCGCG 1406
2118 SerArgSerGlyProArgPro-----ArgArgAlaCysProProArgArgPro 2133
1405 CGAACACACACACGCGCTTGTTCGAGA-----GCACCAAGGC 1370
2134 ArgAlaProGlySerProAlaAlaArgArgArgTrpProArgArgArgProAlaProSer 2153
1369 TCACGAACCTTACGACACACACACGCGCTGGGCGCTTCTTCCCAAGTGGATGA 1310
2154 SerArgArgThrArg-----ProAlaGlyArgGlyArgAlaAlaValSerArgAla 2170
1309 GCATGTTGCTGT-----CCACA 1292
2171 AlaHisCysAlaSerArgSerProGlySerSerAlaArgTrpProAspArgProPro 2190
1291 GACGCGCTGCTCT---CCATTGTGGCG-----CGTTGGCGAGCGCTTGCAGG 1244
2191 SerGlySerAlaProArgArgLeuGlyAlaArgValArgArgTrpProGlySer-----2208
1243 TCTTCCCGCAGTCCAGCACACGCGCAGGCACTGCAGCACCAGC-----ACAGCT 1193
2209 SerCysArgAlaProSerArgArgArgGlySerSerGlyAlaGlyArgGlyArgSerGly 2228
1192 TGCCCATGCGCCCTGTTAGTACAGGCGCTGCTGGAGTACAGAGGCGCTTGGACCA 1133
2229 CysArg---ArgProGlySer-----ArgAlaValArgProGlyGlyAlaCysArgArg 2245
1132 AGTTGCTGCTGTTGGCAGCAGGAAGCGCGCAGCTTCTCTCGTACAGCTCTCTCAATGC 1073
2246 AlaProGlySerProAlaAlaArgProProArgProAlaGlyThrProArgProArgGly 2265
1072 GCGGGTGACACCCCTCGC---CGAACAGCTGCCACACGCTCTCAAACTCCAGCTGCTGA 1016
2266 ArgGlyArgGlyProArgAlaArgAlaGlyArgSerThrArgArgSerAlaAlaArg 2285
1015 TGGCCACCTTCTTACGCTCCAGAGCGCGCTGCGCGGGTTCAGTCCACCGACCA 956
2286 SerArgProGlyArgProProArgSerPro-----GlyGlyAlaAlaGlyArgProArg 2303
955 CCGTCCGCGCCCTCGACACAGCAGGTTTCAGGCGCATTCAGCGCGCTAGTTCAGGGTCA 896
2304 ProGlyArgAlaAlaValAlaHisArgArgGlyArgProAlaArgSerValArgGlySer 2323
895 GCACCGTGTCTTGGGGTTGATCTCCATCATCTCCATATCCGCTGCGGGTCTCCACG 836
2324 ProPro-----ArgProProAlaProValProValArgProAlaProArg 2338
835 ACTGCTGTACAGGAAGTGGGCGGGAACATGGGCGCGCTCCACGCGCTCTCTGTGCA 776
2339 ThrAlaCysAla-----AlaGlyLeuProProAlaProProArg---ProAlaAla 2354
775 GGGCGTGGCCAAAGCTGGGCGCGCAATCCACACGCTAGTAGGGGCGCGCAGCCAGC 716
2355 GlyArg-----AlaSerSerAlaProProArgProProArgArgProGlyVal 2370
715 GCACGTAGGGGATCGAACCTCGGTGTTCTGTCTCCACACGCGCTCCAGCTTCTGTCCA 656
2371 ProArgGlyProProProProAlaArgAlaAla-----ArgArgProValArgTrpSer 2388
655 GGTAGCGCGCGCTCGGGCGCGATGT-----CAATGTTGT-----CGATGCCAAGA 608
2389 GlyAlaArgAlaCysProArgArgCysValaArgArgCysProValGlnProArgArg 2408
607 TCGATCGCCAGA-----AGAAACGCGCGCAGCAGCAGCATCTGCGCGAGGG 563
2409 SerAlaGlyArgArgProProSerGlyArgArgSerAlaAlaArgProArgGlyAlaGly 2428
562 GCAGGTGCTACTTGGCGCTCACTAGAGTTCGGCAACGCCACCGAGCCGT-----CTTGG 506
2429 AlaAlaGlyThr-----SerArgArgArgAlaProGlyArgProSerGly 2443

```

505	QY	ACAGGTACGAGCAGCGCTGCTCGATGACGTTGTGGAAACGGTGGAAATCATCGTGAGCGAGT	446
2444	Db	ThrArgProSerProProPro-----	2451
445	QY	AGGAGAAGGTGATCAGCGTTCGCGGTGCCTCTCAGGGGGCGCAAAATTGGCAAGCGTCGGCCT	386
2452	Db	-----GlyAlaAlaCysProArgGlyProThrAlaGlyProProGlyPro	2466
385	QY	CCA---CGACCTGGA-----CATTCCTCCAGCCCTTGGCTTCGCCTTCTTCTTGG	338
2467	Db	ProAlaArgAlaGlyProValAlaHisGlySerValProGlyProProVal-----	2483
337	QY	CCACCTGGCACGCGAGTGGCACAGGTCGACCGTAGATGGACTTGGCCAGGT	278
2484	Db	-----ArgProArgArg-----	2487
277	QY	CGATGTAATCAGCCATCATATCGACATTCTCCCGAGTGCACACCCAGGTCAACCCAGA	218
2488	Db	-----ArgValArgAspArgProPro-----ProAlaProProArg	2500
217	QY	TGAGGTTTCGAGGCTCGGCCAGCGGGCCACAAAGGCGCGGCTGGGGCCGTAGAGC	158
2501	Db	-----GluAlaAlaArgProGly-----ProProProProAlaHisSer	2513
157	QY	TCCTCCAGGCGAGCAGCGT---GATCATCGC---CCTCTTGTGTCGCGCAACCATATGGC	104
2514	Db	AlaProSerArgProArgProSerHisArgValProAlaSerCys-----	2528
103	QY	GCAGAACGGTACGGTCATCCTTCATGCTGCTGAGCTTGAGCTTCTCCAGGAGGAGTTCT	44
2529	Db	-----HisPro-----GlyGlyGlyArgGluAsp	2536
43	QY	TCCTTGGTGTAGC	32
2537	Db	GlyTrpCysAla	2540

## RESULT 8

```

US-10-084-846A-8
; Sequence 8, Application US/10084846A
; Publication NO. US20040006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLENWEG, AGNES
; APPLICANT: TREFFZER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084, 846A
; CURRENT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 8
; LENGTH: 19608
; TYPE: PRT
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: Protein 3: amino acid sequence encoded by coding strand 1.
; OTHER INFORMATION: Start codon: atc, Start position: nucleotide 3.
US-10-084-846A-8

```

Alignment Scores:	1.23e-18	Length:	13608
Pred. No.:	Score:	Matches:	205
	446.00	Conservative:	57
	Percent Similarity:	Mismatches:	248
	37.5%	Indels:	188
	Best Local Similarity:	Gaps:	34
	29.4%		
	Query Match:		
	12.3%		
	DB:		
	4		

US-10-620-914-44 (1-1947) x US-10-084-846A-8: (1-19608)

Qy	17	ACGCCCGGC-----CTGCCAGCTACACCAAGAAGAACTTCTCCCTGGAGAACTCA	67
Db	5039	ThrAlaGlyArgThrProLeuArgArgAlaProArgGlyThrHisPro-----	5054
Qy	68	AGCTCAGCAGCATGAAGATGACCTGACCTGCTGGCCCATATGTGTTCGGCAGCAAGA	127
Db	5055	-----AlaSerGlySerArgArgArg	5061
Qy	128	AGGGCGATGATCAGCTGCTCGCTCGAGAGCTTCTACGGGCCCCAGGCGCTGCTCTTG	187
Db	5062	ArgSerSerGlnValArgLeuGlyGlnTyrAlaGlySerHisAlaPro-----	5078
Qy	188	CTGCCCGCTGGCCGAGCGCTCAACCTCATCTGGGTTGACCTGGGTGTGGCA----	241
Db	5079	LeuGlnAspGlyArgGlyAlaArgHisSerArgSerGlyCysTyrIleValaHisGlu	5098
Qy	242	-----CTGGGGAGAAATGTGCAATATGATGGCTG	268
Db	5099	GluGlyLeuGluileProLeuGluArgArgAspTyrThrArgArgSerIleSerGlnLeu	5118
Qy	269	-----ATTACATCGACCTGG	283
Db	5119	ProValProAspValHisPheTyrProGlnCysAlaAlaThrAsnLeuAsnTyrGlnTyr	5138
Qy	284	CGAAGTTCAAAGTCCATCTACGTGGTGTGCACCTGTGCCACTGTGTCGAGGTGGCCAGA	343
Db	5139	GlyAlaProCysProIysGlyCysSerSerAlaValGlyArg-----ArgAspProArg	5156
Qy	344	AGAAGCGAAGCCAGGCGCTGGAAGAATGTCCAGGTGGTG-----AGGCCGACGCTT	397
Db	5157	ArgProProCysSerAlaArgProAlaCysProSerTyrTipAsnGlyArgProSerArg	5176
Qy	398	GCCAAATTTGGCCCCCTGAGGCACCGGCACGCTCATCCTTCTCTACTCGCTCAGA	457
Db	5177	AlaThrThrSerAlaSerArgSerProProArgAlaAlaProSerSerThrSerSerAla	5196
Qy	458	TGATTCACCGTTCCACAACGTCACTGACACAGGCTTGCTGCTACTGTGCCAAGACGGCC	517
Db	5197	ArgTyrArgArgSerThrProAlaValThrArgSerArgThrAlaCys-----	5213
Qy	518	TGTTGGCGTTCGCACTTCTACGTGAGCGGCAAGTACGACTGCCCTCGCCAGATGTC	577
Db	5213	-----CysCysAlaGlyAlaArgArg	5219
Qy	578	CTGTTGCGCGCGTTCCTCTGCGCATCGATCTTCGACATCGACAACATTGACATCGGCC	637
Db	5220	ThrGlyArg-----SerThrGlyArgArgSerSerValPro-----ThrSerAla	5234
Qy	638	CCG-----AGCGCGCGCTACCTGGAGCAGAAGCTGGAGCGCGTGTGGAGCAGACA	691
Db	5235	ProGlyArgSerThrAlaThr-----SerThrThrCysSerArgThr	5249
Qy	692	CCAGGTTTCGATCCCTACGTGCGCTGCTGCGGCCCTTACTACGTGTGGATGTGCC	751
Db	5250	ArgSerSerArgAlaProThr-----SerProArgAlaSerArgSerSerAla	5265
Qy	752	GCTTGCCACAGCTTGGCCACGCCCTGCACGAGGACGGGTGGAGCGCGGCCCATGTGCC	811
Db	5266	CysCysSerThrAlaThrAlaArgLeuGlyArgSerGlyTyrThrProThrProValSer	5285
Qy	812	CGCCACCTTCTGTACACGCAAGTCTGGGAGGACCCCGAGCCGATATGGAGGTGATGG	871
Db	5286	-----TyrSerAlaArgSerThrThrSer-----	5294
Qy	872	AGATCAACCCCAAGCACAGTGTGTCACCTGACTAGCGGGCGGTGCAATGCTCTGAACC	931
Db	5295	---SerThrProArgAla-----	5299
Qy	932	TGCTGTGACGGGGCGCGCCAGTGTGTGTGCTGATCGATCGACATGCCCGGCGATCGCGC	991
Db	5300	-----GlyProAla-----SerSerProSerThrArgThr	5309

```
QY 992 TTCTGGAGCTCAAGAGGTGGCCATTGAGGAGTGGAGTGGAGAGCTGTGGCAGCTGT 1051
Db 5310 -----AlaAlaArgThrArgSerSerAla 5317
QY 1052 TCGCGCAGG-----GGCTGCACCGCGGCTGAGAGCTGTACGAGAAGAGCTGGCGC 1105
Db 5318 ThrSerArgSerGlyAlaThrGlyArgAlaAlaArgCysProValArgProGlnGly 5337
QY 1106 CTTTCTGTGTCGAACACGACCAACTCTGTGTCCA-----AGGCGCTCTGGT 1153
Db 5338 AlaSerThrSerSerAlaProArgThrAlaGlyThrGlySerArgCysAlaGlyThrGly 5357
QY 1154 ACTTCCAGCAGCGCTGTACTACAGGCGGCATGGCAGCTGTGTGGTGTGGCAGT 1213
Db 5358 ThrAlaSerValSerAlaThrArgThrGlySerThrProAlaAlaGlyAsnThrPro 5377
QY 1214 GCCTGGCGCTGTGTGGGAC---TGGGCAAGACCGTCAAGCGCTCGCCAACGCGCCA 1270
Db 5378 ThrTrpArgArgCysTrpProArgTrpSerArg-----AsnProArgProCysAlaThr 5395
QY 1271 CAATGGAGGACGCGCGTGTGGGACGACCAATGCTCATCTTCCTGTGGAGAACG 1330
Db 5396 SerTrpArgThrAlaSerThrSerArgAlaTrpGlyTrpSerArgThrSer---ProThr 5414
QY 1331 GCGCCAGACCGCTGTGTGGTGTCTCAAGTTCGTGAGCTGTG-----TGCTCTTCA 1384
Db 5415 SerProThrAlaSerThrAlaProAlaThrThrSerProAlaThrProProAlaSerSer 5434
QY 1385 ACAAGGCGGTGTGTGGTGTGGCGGCGGTGGCGGCAAGCAGTACGCGTGTCAAGG 1444
Db 5435 ThrArg---CysCysProProAlaSerThrTrpArgSerThrAlaAlaCysTrpProArg 5453
QY 1445 CGGAGCGCATCCCATTTGAGAACTATATCGCGCGCACCATGACGCGCTGGCGGAGAACT 1504
Db 5454 ArgProCysSerProArgThrArgThrSerArg-----5464
QY 1505 GCGAGTGGCGGACGAGAACTACTTCTACTACACTGCTCACCGGCAAGTTCCTCGCGG 1564
Db 5465 ArgArgArgProGlyArgSerThrSerArgSerThrAlaThrProThrSerGlyCysSer 5484
QY 1565 ACAACTGCCACCTACCTACGCGAGGCGGCTTCGCCACCTCAAGAGTGGCGGTGG 1624
Db 5485 ProTrpCysProAlaCysThrSerAsnArgProAlaArgProProThrSerSerTrpPro 5504
QY 1625 ACA-----ACCTGACCGTCTCCCAACTTCTCATGAGAGAGCTCAAGCGCGCA 1675
Db 5505 ThrArgProArgSerAlaProArgProSerThrSer-----ArgSerThrAlaArgTrp 5522
QY 1676 CTTACACAGAGTGATTCGTATGGACCGAGCTGGCTGGATATGCCCGTGGCCAAG 1735
Db 5523 ProSerPro-----AsnTrpSerArgAspTrpProThr 5533
QY 1736 AGC---TGGCGAGTGCGTGGCCAAAGAGTGGCGCGCGGCGCATCG---TCATCTGGC 1789
Db 5534 SerThrThrProSer-----ProAlaArgThrArgThrAlaProSerSerArgThrPro 5551
QY 1790 GCTCGCGCTCCCTCAGCCCGCTCAGCGAGCTGATCC-----1828
Db 5552 ArgProThrProCysArgArgThrThrProSerAlaSerCysSerArgArgProArgThr 5571
QY 1829 -----AGAGGGGGGCTTCGAGTGGCGTGTGATCGCGCGCGGCTACAGGCTACATGG 1882
Db 5572 ProValTrpArgArgProGlyCysProProHisArgSerAlaAlaGlyProAlaArgTrp 5591
QY 1883 ACGCGCTCAACATGTACAGCTCT-----TCTACATGGCGCGCGGAGG 1927
Db 5592 ThrProThrThrSerThrThrProProArgValSerThrTrpProAlaArgArg 5609
```

## RESULT 9

US-10-084-846A-3

; Sequence 3, Application US/10084846A

; Publication No. US20040006026A1

```
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLNENWEG, AGNES
; APPLICANT: TREFZER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084, 846A
; CURRENT FILING DATE: 2003-02-25
; PCT/P01/09815
; PRIOR APPLICATION NUMBER: 2001-08-24
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 3
; LENGTH: 19695
; TYPE: PRT
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: Protein 1: amino acid sequence encoded by coding strand 1.
; OTHER INFORMATION: Start codon: gga, Start position: nucleotide 1.
; US-10-084-846A-3
```

## Alignment Scores:

Pred. No.:	1,75e-18	Length:	19695
Scores:	443.50	Matches:	229
Percent Similarity:	34.3%	Conservative:	48
Best Local Similarity:	28.4%	Mismatches:	213
Query Match:	12.3%	Indels:	317
DB:	4	Gaps:	51

US-10-620-914-44 (1-1947) x US-10-084-846A-3 (1-19695)

```
QY 23 GGCCTGCGAGTACA-----CCAAGAGA-----ACTTCTCCTCGAGAGCTCAAGC 70
Db 8455 GlyGluArgArgThrHisProProArgArgHisTrpThrGlyProArgSerSerGly 8474
QY 71 TCAGCAGCATGAAGATGACCTGACCTGCGCATATGTGGTTCGCGCAGCAAGAGG 130
Db 8475 SerAsp-----GlyThrAlaSerArgThr 8482
QY 131 CGCATGATCAGCTGCTCGCTGGAGAGCTTCTACGGGCCCCAGG-----CGCTG 181
Db 8483 -----ArgAlaArgPheSerProArgThrArgSerProVal 8494
QY 182 -----CCTTGTGCGCGCTGCGCGAGCTCAAGC 214
Db 8495 SerArgArgAlaProAsnGlySerSerProAlaSerProThrArgAlaCysAlaProThr 8514
QY 215 TCATCTGGTGTGACCTGGTGGCTGCGACTGGGAGAAATGTCGATATGATGCTGATTACA 274
Db 8515 ThrThrGly-----Thr 8518
QY 275 TCGACCTGCGCAAGTTCAAGT---CCATCTACGTGGTGCAGCTGTGCCACTCGCTGTGG 331
Db 8519 SerProSerArgAlaMetSerValProSer-----SerThrAlaCysThr 8533
QY 332 AGTGGCCAAAGAAAGGAGGCAAGGCTGGAAGA-----ATGTCCAGGTGCTGG 385
Db 8534 ThrTrp---ArgSerArgThrGlyProArgAsnGlyThrCysCysThrAlaArgAsnTrp 8552
QY 386 AGGCGCAGCTTGCCAAATTTGGCCCCCTGAGGCGACCGCAGCTCATCCTCTCTCT 445
Db 8553 ArgSerSer-----ProProArgSerTrpThrAlaArgSerCysProProPro 8568
QY 446 ACTCGCTCA-----CGATGATTCACCGTTCCAAACGTC 481
Db 8569 ThrArgSerSerSerArgSerArgThrGlyProArgCysHisGlyThrValThrAla 8588
QY 482 TCGACCGAGCTT---GCTGTACCTGTCCAAAGCGGCTGGTGGCGGTTCGCGACTTCT 538
Db 8589 SerThrSerAlaArgAlaArgSerAlaIleAlaSerAlaTrp-----8602
```



QY	539	ACGTGAGCGGCAAGTACACACTCGCCCTCGCCACAGATGCCCTGGTCGCGCGGTTCTTCT	598
DB	8603	-----ThrThr-----ProAspArgThrThrAspAla	8611
QY	599	-----GGCGATCGATCTTCGACATCGACACATTTGACATCGCGCCCGGAGCGCC	646
DB	8612	TrpArgArgPheArgAspArgThrCysSerProThrThrProArgTrpArgArgCysAla	8631
QY	647	CGCGCT---ACCTGGACAGAACTGGAGCGCG	679
DB	8632	ArgProAspProTrpPheArgSerArgSerAlaArgAlaThrTrpTrpCysThrThrSer	8651
QY	680	GGGAGCAGACACCCCGGTTGATCCCTACGTCGCGTGC	721
DB	8652	GlySerSerThrValArgAlaArgThrProThrAlaValGlyGlyProSerSerSer	8671
QY	721	-----	721
DB	8672	SerThrProThrArgProHisArgArgArgProGluArgCysAlaAlaGluArgGlnAla	8691
QY	722	-----TGCGGCGCCCT	733
DB	8692	AspIleGluGlnValArgGluSerPheGlyLeuCysSerProProArgThrGly	8711
QY	734	-----ACTACGCTGGATTGGCGCCTGCCACCGTTGCCACGCCC	784
DB	8712	ArgAspIleThrCysAlaTrpCysProTrpProGlyArgCysAlaProArgAspThrArg	8731
QY	785	AGCGCGTGG-----AGCGCGCGCCCATGTTCCCGCCCA	817
DB	8732	SerGlyTrpArgAlaArgArgArgAsnAlaGlySerProArgProAlaCysArgSer	8751
QY	818	CTTCTCTGTACAGCACTCTGTGGAGACCCCGACCGCGATATGGAGTGATGGAGATCA	877
DB	8752	ArgCysSerThrArgArgThrTrpLysAlaProGly-----TrpArgSer	8766
QY	878	ACCCAAAGGACACGGTCTGCTACCTGACTAGCGCGGCTGCAATGCCCTGAACCTGCTGG	937
DB	8767	-----MetCysArgPro-----	8770
QY	938	TGC-----AGGGGCGCGCCAGGTGG---TGCTGGTGGAGTGCACACCCGCGCAGT	985
DB	8771	CysThrHisArgArgSerProValArgGlyArgCysArgCysThrArgSerProGlySer	8790
QY	986	CGGCGCTTCGGAGCTAGAGAGGTGCCATTACGACGCTGGAGTTTGAGGACGTGTGC	1045
DB	8791	Arg-----CysSerArgSerThrThr	8797
QY	1046	AGCTGTTTCGGAGGGGTGCACCCGCGCATTCAGGAGCTGTACGAGAGAACGCTGGCGC	1105
DB	8798	SerThrProAlaThrCysAlaThrSerGlyArgSerArgSerThrArgCysSerAlaAla	8817
QY	1106	CTTCTCTGTCCAAACCCAGCACCACTTCTGTGTCCAAAGCGCCTCTGGTACTTCCAGCAGC	1165
DB	8818	MetThrThrArgAlaSerAlaThrThrGlyPro-----ThrTrpTrpCysThrThr	8835
QY	1166	GCCTGTACTACGAGCGGCATCGGCAGAGCTGTGCTGGTGTGCTGAGTGCCTGGCGCTGG	1225
DB	8836	SerTrpProSerArgAlaHisTrpSerProArgSerGlyAlaCysArgAla-----	8852
QY	1226	TGCTGGAGCTGGGCAGACCGCTCAAGCGCTCCGCAACGCGCCCAAA-----	1273
DB	8853	-----SerMetCysArgProGlySerSerAlaProTrpArgProSerArgGlySerThr	8870
QY	1274	TGG-----AGGAGCAGCGCGCTGTGTGGGACAGCAACA	1306
DB	8871	TrpSerArgArgThrArgCysArgAlaSerArgSerThrAlaSerGlyAlaAlaThr	8890
QY	1307	TGCTCATCCACTTCGTGAGAACGGGCCCA-----AGCGCTGG---	1345
DB	8891	GlySerSerThr-----ArgTrpThrProHisArgThrTrpArgCysArgArgTrpAla	8908

1346 QY -----TGCGCTGTTGCTCAAGTTCGTGAGCCTCGTCTCTTCAACAAGCGCG----- 1399

8909 Db ThrArgCysGlyCysProCys-----AlaThrCysProThrThrAlaProArgAla 8925

1394 QY -----TGCTGTGGTTCGGCGCGCGCGGTGC 1417

8926 Db ArgThrArgGlySerTrpGlyArgSerGlyAlaAspAlaCysAlaSerSerGlyAlaThr 8945

1418 QY -----CGGCAAGCAGTACGCGCTGATCAAGCGGCGGCTCCCA---TTGGAACCT 1468

8946 Db ArgProArgAlaSerSerAlaProThrTyArgArgCysGlyThrProSerThrArgArg 8965

1469 QY ACATCGCGCGCA-----CCATGGACGGCGTGGCGG 1498

8966 Db CysGlyArgAlaArgArgTrpCysSerArgArgArgAsnArgTrpArgSerTrpGlu 8985

1499 QY AGAACTCGGACGTGGCGCAAGCAGAACTACTTCTACTACAACGCTCCCTCAGCGGCAAGTCC 1558

8986 Db ArgCysArg-----ProAlaSerGly 8992

1559 QY TGGCGGACAACTGCCCCACCTACTCGCGAGGCGCCTTCGCCACCCCTCAAGAGTGGCG 1618

8993 Db CysCyAlaThrAla-----ArgTrpSerSerSerSerArgThrAla 9006

1619 QY -----TGTGGACAAC---TGACCGTCTCCACCACCACTTCTTCATGGAGGAGCTCAAAG 1669

9007 Db ThrCysTrpCysThrThrAlaAlaProThrAlaThrThrAlaSerTrpArgGlyPheArg 9026

1670 QY CGGCACTACACCAAGGTGATTCTGATGGACCAAGTGGACTGGCTGGATATGCCCGTGG 1729

9027 Db SerCysProSerProSer-----ThrThr----- 9034

1730 QY CCAACGAGTGGCGGAGTGCTCGGCCAAGCAGAGTTGCGCGCGGCGCATCGTCATCTGGC 1789

9035 Db -----ThrArgSerAlaAlaValSerIleProPro 9045

1790 QY GCTCGCGCT-----CCCTCAGCGCGCTTACCGCGAGCTGATCCAGAAGCGCGGCTTCG 1843

9046 Db AlaProProCysProAlaLeuArgProArgProArgSerArgArgArgCysAla 9065

1844 QY ACGTGC-----GCTCATCCGCGCGCCACTCAGGCTCATGGACCGCGCTCAACA 1894

9066 Db GlyCysCysSerThrThrAlaThrAlaGlyArgArgAsnGlyThrAlaSerSer 9085

1895 QY -----TGTCAGCTCTTCTACATGTCGCGCGCGCGCA 1924

9086 Db GlyHisLeuArgProProArgSerProGlyCys-----ProAlaSerTrpArgThrAla 9103

1925 QY AGGCGGCCCAAGAGGACAACT 1945

9104 Db GlyCysArgArgArgThrSer 9110

RESULT 10

US-10-084-846A-8

Sequence 8, Application US/10084846A

Publication No. US2004006026A1

GENERAL INFORMATION:

APPLICANT: WEITNAUER, GABRIELE

APPLICANT: MOHLENWEG, AGNES

APPLICANT: BECHTOLD, AXEL

APPLICANT: BECHTOLD, ANDREAS

TITLE OF INVENTION: AVILAMYCIN DERIVATIVES

FILE REFERENCE: 1974-005

CURRENT APPLICATION NUMBER: US/10/084,846A

CURRENT FILING DATE: 2003-02-25

PRIOR APPLICATION NUMBER: PCT/EP01/09815

PRIOR FILING DATE: 2001-08-24

PRIOR APPLICATION NUMBER: DE 101 09 166.4

PRIOR FILING DATE: 2001-02-25

NUMBER OF SEQ ID NOS: 120

SOFTWARE: PatentIn Ver. 3.2

SEQ ID NO 8

LENGTH: 19608

; TYPE: PRT  
 ; ORGANISM: Streptomyces viridochromogenes  
 ; FEATURE:  
 ; OTHER INFORMATION: Protein 3: amino acid sequence encoded by coding strand 1.  
 ; OTHER INFORMATION: Start codon: atc, Start position: nucleotide 3.  
 US-10-084-846A-8

Alignment Scores:  
 Pred. No.: 5,05e-18 Length: 19608  
 Score: 436.00 Matches: 217  
 Percent Similarity: 35.0% Conservative: 40  
 Best Local Similarity: 29.6% Mismatches: 237  
 Query Match: 12.0% Indels: 240  
 DB: 4 Gaps: 44

US-10-620-914-44 (1-1947) x US-10-084-846A-8 (1-19608)

```

QY 1939 CCTTCTTGGCGCCTTCGGCGGCCATGTAGAGGAGCTGTACATGTTGACGGCTCCA 1880
DB 10543 ProArgThrArg-----CysProArg----- 10549
QY 1879 TGTAGCCTAGTGGCGGGGATGCGACGCGACGTGGAAGCCCGCTTCTGATACAGCT 1820
DB 10550 -----ProArgArgProSerProThrAlaArgProProArgGlyGlySerAla 10567
QY 1819 CGGCGTAGGGCGGCTGAGGAGGGGAGCGCCAGATGACGATGCCGCCCGCGCAACCT 1760
DB 10568 AlaArgArgPro-----ArgArgAlaValProAlaAlaValProAlaValSer 10585
QY 1759 GCTTGGCCAGGCACTCGGCCAGCTGTTGGCCAGCGGATATCCAGGCGATCCACGTGGT 1700
DB 10586 ProProProGlyGlySerGlyGlyArg-----ThrAlaProGlyAlaAlaProProArg 10603
QY 1699 CCATCAGAATCACCTTGGTGTAGTGGCGCTTTGAGCTCTCCATGATGAAGAAGTTGGTG 1640
DB 10604 ProSerArg---ProTriProGln----- 10610
QY 1639 AGACGGTCAGGTTGTCCAGCGCCACTCTTGAGGGTGGGAGGCCCTTCGC---CCA 1583
DB 10611 GlyArgThrGlySer-----GlyTriArgArgArgAspArgTyGly 10624
QY 1582 GGTAGTGGGCGAGTGTGCGCAGGAACCTTCCCGGTGAGGCGAGTTGTAGTAGAGTAGT 1523
DB 10625 GlyArgArgSerAlaCysArgSerGly-----ArgGlyThr 10636
QY 1522 TCTGCTTGGCAGCGTGGAGTTCTCCGCCAGCGCTCCATGGTGGCGCGATGTAGTTCT 1463
DB 10637 Ala-----AlaGluSerProProArg-----CysArgSerCysSerSer 10649
QY 1462 CAATGGGATGCGCTCCGCTTGATCAGCGCGTACT-----GCT 1424
DB 10650 ---TriArgCysArgGlyAlaProAlaGlyArgSerThrGlyGlyAlaProAlaGlySer 10668
QY 1423 TSCCGGCGAGCGCGCGCGGACACAGCAGCGCTTGTGAGAGCAGCAGCGCTCACGA 1364
DB 10669 CysAlaAlaProGlyArgArgProAlaAlaLeuPro-----AlaPro----- 10682
QY 1363 ACTTGACGAACCAACACACAGCGGCTTGGCGCGCTTCTTCCAGAAAGTGATGAGCATGT 1304
DB 10683 -----ProProValProAlaAlaProAlaPro----- 10691
QY 1303 TGCTGTCCACAGCGCGCTGCTCTCTCCATGTGGGCGGTTGGGAGCGCT----- 1250
DB 10692 -----ProProArgProAlaAlaProPro-----ArgArgTriArgAlaAlaAlaPro 10707
QY 1249 -----TGACGGTCTTCCAGTCCAGCA-----CCACGG----- 1220
DB 10708 ValArgArgArgArgCysProArgProSerCysProArgArgCysThrArgValCys 10727
QY 1219 CCAGGCACTGCACACCCAGCAGCTTGCCTCCATGCGCCCTGGTAGTACAGCGCGTGT 1160
DB 10728 ProGly---SerAlaProAlaAlaThrGlyPro---ArgProGlyArgArgCysArgAla 10745

```

```

QY 1159 GGAAGTACCAGAGCGCTTGGACCAGAAAGTTGTGGCTGGTTTTCGACAGAAAGGCGCCCA 1100
DB 10746 ArgSer-----AlaTriAlaAlaAlaAlaAlaArgPro 10756
QY 1099 GCTTCTTCTCGTACAGCTCCTCAATGGCGGGTGCACGCCCTCCGCAACAGCTGCCACA 1040
DB 10757 Ala-----AlaAlaProProArgProValArgGly 10766
QY 1039 CGTCTCAAACTCAGCTGCTGAATGCCACCTTCTTTCAGCTCCAGAAAGCGCCAGCTCG 980
DB 10767 ProProArgThrGlySerGlyArgGlyProThrArgArgArgThrSerArgSerThr 10786
QY 979 CGGGTTGTCAGTCCACCGCACACCTGCGCGGCCCTCCACACAGAGTTTCAGGGCAT 920
DB 10787 AlaGly-----ProProArgProGlyArg-----AlaProAlaArgSerProArg 10801
QY 919 TGCAGCGCGCTAGTACGGGTACGACCGTGT---CCTTGGGGTTGATCTCCA---TCA 866
DB 10802 CysArgThrArgAlaThrHisArgArgSerAlaArgProGlyAlaAlaArgProValArg 10821
QY 865 CTTCCATATCCGGCTCGGGTCTCCACGACTCGGTGTACAGAAAGTGGCGCGGAACA 806
DB 10822 ProArgArgProCysArgAlaProProGlyArgAlaAlaSerGlyArgArgSerPro 10841
QY 805 TGGCGCGCGCTCCACGCGCTCTCTGTCAGGGCGCTGCCAACCTGGGCGAGCGGCA 746
DB 10842 ArgCysHisSerCysAlaAlaProArgSerThrSerSerArgSerArgAlaGlyArg 10861
QY 745 TCCACAGCTAGTAGGGCGCGGAGCCAGCAGCAGTAGGGATCGAACCTTGGGTGTCT 686
DB 10862 ArgArgArgSerArgAlaThrAlaSerAlaAlaArgThrGlyAlaAlaAlaAlaAla 10881
QY 685 GCTCCACACGCGCTCCAGCTTCTGCTCAGGTAGGCGCGCGCTCGGGCGCGATGTCAA 626
DB 10882 AspArgThrAlaAlaValCysThrTriProAlaHisArgThrProAlaGlyArgPhePro 10901
QY 625 TGTGTGATGTCGAAGATCGATCGCCAGAAAGAACCGCGCAGCAGG----- 578
DB 10902 SerCysProAlaAlaArgAlaProGlyProAlaSerGlyArgSerArgArgProArgVal 10921
QY 577 -----GCATCTGGCGCA 566
DB 10922 ProAlaThrAlaGlyThrProProGlySerArgValProArgArgAlaAlaProArgPro 10941
QY 565 GGGCAGGTCTGCTTGTCCG-----TCACGTAGA-----AGTCGGCAAGCG 524
DB 10942 GlyAlaProCysThrCysAsnValProProSerSerArgArgThrSerSerSerGluArg 10961
QY 523 CCACCA-----GGCGCTCTTGGGACAGGT 500
DB 10962 LeuProAspLeuProPheAspValCysLeuAlaLeuGlyGlyAlaAlaAlaAlaPro 10981
QY 499 ACGCAAGCCTGTGTCGATGAGCTTGTGGAACGGTGGATCATCTGAGCGAGTAGGAGA 440
DB 10982 AlaValArgProGlyArg-----ArgThrArg 10990
QY 439 AGGTGATGAGCTCGCGTGC-----CCTCAGGGCGCGCAATTCGCAAG 395
DB 10991 ArgTriSerAlaAlaAsnCysArgTriArgSerGlyProAsnArgGlyArgAlaGlyArg 11010
QY 394 CGTGGCGCTCCA---CGACCTGGACATTTCTTCCAGCCCTTGGCTTCGCT----- 347
DB 11011 ArgAlaProProArgArgProGly-----ProGlyProGluProArgValArg 11026
QY 346 -----TCTTCTTGGCCACTCGCACAGGAGTGGCACAGGT----- 311
DB 11027 SerArgAlaProSerProProArgArgArgArgGlyAlaGlyAlaAlaProGluProPro 11046
QY 310 -----CGACCACTAGATGACT----- 293
DB 11047 ProGlyIleArgCysAlaValArgArgArgProGlyArgCysSerAspCysArgProCys 11066
QY 292 -----TGAACCTTCGCCAGGT 278

```

Db 11067 AlaGlyArgArgArgGlyHisGlyAlaMetAlaProProArgProGlyThrAlaProGly 11086  
 QY 277 CGATGTAATCAGCCATCATATCGCATCTTCCCGAGTCCACCCAGGTCACCCAGCA 218  
 Db 11087 ArgAlaArgArgArgTIPAlaArg-----ProGlyArgProArg 11099  
 QY 217 TGAGGTCGAGCGCTCGG-----CCAGCGGGCAGAAAGCGAGCGCCCTGGGGCCCGT 164  
 Db 11100 ThrArgArgArgAlaAlaProValProArgGlyAlaAlaGlyProValProGlyProSer 11119  
 QY 163 AGAGCTCTCCA-----GGCGAGCAGCGTGAT-----CATCGCCCT----- 128  
 Db 11120 ProAlaSerProSerCysAlaArgGlyArgGlyArgAspAlaHisArgProArgArgGly 11139  
 QY 127 ---TCTTGCTGCCGAACCATATATGGCGCAGAACCGTCAAGT 89  
 Db 11140 SerSerSerArgArgSerAlaGlyProArgArgSerGly 11153

## RESULT 11

US-10-084-846A-4  
 ; Sequence 4, Application US/10084846A  
 ; Publication No. US20040006026A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: WEITNAUER, GABRIELE  
 ; APPLICANT: MUHLERWEG, AGNES  
 ; APPLICANT: TREFFZER, AXEL  
 ; APPLICANT: BECHTHOLD, ANDREAS  
 ; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES  
 ; FILE REFERENCE: 1974-005  
 ; CURRENT APPLICATION NUMBER: US/10/084,846A  
 ; CURRENT FILING DATE: 2003-02-25  
 ; PRIOR APPLICATION NUMBER: PCT/EP01/09815  
 ; PRIOR FILING DATE: 2001-08-24  
 ; PRIOR APPLICATION NUMBER: DE 101 09 166.4  
 ; PRIOR FILING DATE: 2001-02-25  
 ; NUMBER OF SEQ ID NOS: 120  
 ; SOFTWARE: PatentIn ver. 3.2  
 ; SEQ ID NO 4  
 ; LENGTH: 19725  
 ; TYPE: PRT  
 ; ORGANISM: Streptomyces viridochromogenes  
 ; FEATURE:  
 ; OTHER INFORMATION: Protein 2: amino acid sequence encoded by coding strand 1.  
 ; OTHER INFORMATION: Start codon: gat, Start position: nucleotide 2.  
 US-10-084-846A-4

Alignment Scores:  
 Pred. No.: 2,77e-17 Length: 19725  
 Score: 424.00 Matches: 222  
 Percent Similarity: 33.2% Conservative: 41  
 Best Local Similarity: 28.0% Mismatches: 251  
 Query Match: 11.7% Indels: 278  
 DB: 4 Gaps: 49

US-10-620-914-44 (1-1947) x US-10-084-846A-4 (1-19725)

QY 1939 CCTCTTTGGCGCCCTTCGGCGGGCCATGTAGAAGGAGCTGTACATGTGTGACGGGTCCA 1880  
 Db 17234 ProSer---ArgProTyrGlyArgProTTPAsnArgSerMetSerAsnTTPProGly--- 17251  
 QY 1879 TGTAGCCCTGAGTGGCGCGG-----GGATGCAGCGCAGCGTTCGAAGCCCGCCT 1832  
 Db 17252 -----AlaSerTTPArgSerSerArgTTPProGlyCys-----ArgThrSerProArg 17267  
 QY 1931 TCTGGATCAGCT-----CGCGGTAGGCGGGCTGAGGAGCGGAGC----- 1790  
 Db 17268 AlaGlyCysAlaTTPGlyArgTTPArgArgAlaGlySerGlyArgSerAlaArgCysSer 17287  
 QY 1789 -----GCCAGATGACGATGCGCGCCGCAACCTCTTGGCCAGGCACTTCGGSCCA--- 1739  
 Db 17288 ThrProAlaArgThrGlyCys-----AlaSerSerAlaGlySerGlySerArgProPro 17305

QY 1738 ---GCTCGTTGG-----CCAGCGGCATAT---CCAGCCAGTCCA--- 1706  
 Db 17306 ThrArgArgTTPProProProSerGlySerProArgLeuTyrGlyThrAlaSerProTyr 17325  
 QY 1705 ---CGTGGTCCATCAGATCAGCT---TCG-----TGTAGGTCG 1673  
 Db 17326 ThrSerThrTTPThrGlyProSerProGlyProProHisTTPSerSerArgCysThrCys 17345  
 QY 1672 GCGCTTTGAGCTCCTCCATGAAGAAGTTGGTGGAGACCGTTCAGGTTCCTCCACCGCCAC 1613  
 Db 17346 -----TTPTPArgArgAla-----ProArgHis 17353  
 QY 1612 TCTTAGGGTGGCGAAGCCG-----CCTCGCGCA 1583  
 Db 17354 AlaGlyProSerArgArgAlaGlyArgArgSerSerAlaArgCysTTPProArgArg 17373  
 QY 1582 GGTAGTGGGCGAGTGTTCGCGCAGGAAGTTCGCGGTGAGGAGTGTGTAGTAGA----- 1529  
 Db 17374 GlyArg-----ArgGlnAlaGlyCys-----AlaCysCysSerArgGlyGly 17387  
 QY 1528 -----AGT 1526  
 Db 17388 ProHisGlyAspGlyArgTTPArgProTTPArgArgSerAlaHisProSerArgProThr 17407  
 QY 1525 AGTTCTGCTTCGCGCAGCGAGTTCCTCCGCCACCGCGTCCATGGTGGCGCGATGTAGT 1466  
 Db 17408 ThrThrValCysHisArgCysSerProArgProArgSerThrCysAlaAlaArgAla 17427  
 QY 1465 TCTCAATGG-----GGATGCGCTCGCTTCATCAGCGGCTACTGCT 1424  
 Db 17428 ProArgTTPProProTTPSerThrSerSerCysAlaProPro-----SerThrThrAla 17445  
 QY 1423 TGCCCGCGCAGCGCCGCCGAAACACACAGCAGCGGCTTGTTCGAAGAGCACCA----- 1373  
 Db 17446 CysSerProSerSerArg-----AlaAlaProProThrProArgArgProArgGlySer 17463  
 QY 1372 -----GGTCTCAGAACTTGACGAACAGCACACACCG 1340  
 Db 17464 CysArgThrArgTTPProSerGlySerGlyProArgArgSerThrThrArgThrAlaGly 17483  
 QY 1339 GCTTGGCGCGTTCCTTCCAGNAAGTGGATGAGCATGTTGCTCTCCACAGCGGCTGCT 1280  
 Db 17484 SerHisCysAlaAlaThrArgProSer---AlaProAlaSerProSerProSerAlaGly 17502  
 QY 1279 CCTCATTGTGGCGCGCTTCGCGCAGCGCTTGAGGCTTTCGCCAGCTCCACCGACCG 1220  
 Db 17503 ArgSerSerAlaThrArgTTPGlySerAlaArgArgSerArgProSerProHisSerArg 17522  
 QY 1219 -----CCAGGC 1214  
 Db 17523 ThrProProArgGlyArgAlaThrSerTTPSerValArgProAlaCysSerThrGly 17542  
 QY 1213 ACTCAGCACCCAGCACAGCTTCCCATCGCGCTGTAGTAGTACAGGCGGTGCTGGAAGT 1154  
 Db 17543 ProGlyArgProArgSerAlaValCysCysAlaHisArgCysThrAlaProThrGlyArg 17562  
 QY 1153 ACCAGAGCGCTTGGACCAAGTGTGGCTTTCGCGAAGGAGCGGCCAGCTTCT 1094  
 Db 17563 ThrArgSer-----ProThrGlyArgGlyGlyAlaAla 17572  
 QY 1093 TCTGTACAGCTCCTCAATCGCGGGGTGCAGCGCTCCGCCAAGAGCTGCC----- 1043  
 Db 17574 SerArgSerProSerThrArgCysGlyAlaSerArgProGlyAlaAlaArgTTP 17593  
 QY 1042 -----ACAGCTCTCAAACT-----CCAGTGTGTGAATGCCA 1010  
 Db 17594 AlaCysSerSerTTPThrArgProThrThrArgfileProGlyProAlaAlaProTTPPro 17613  
 QY 1009 CCTTCTTCAGCTCCAGAACCGCGACTCGCGGGTTCAGTCCACCGACACACCTGGC 950  
 Db 17614 SerProSerGly-----ArgSerThrAlaThrAlaCysCysSerProValArgProTTP 1763  
 QY 949 CGGCCCTTCACACAGCAGGTTTCAGGCGCATTCGAGCGCGCTAGTCAGGCTCAGCACCG 890

Sequence 4, Application US/10084846A  
Publication No. US2004006026A1  
GENERAL INFORMATION:  
APPLICANT: WEINHAUER, GABRIELE  
APPLICANT: MUHLNBERG, AGNES  
APPLICANT: TREFFER, AXEL  
APPLICANT: BECHTOLD, ANDREAS  
TITLE OF INVENTION: AVILAMYCIN DERIVATIVES  
FILE REFERENCE: 1974-005  
CURRENT APPLICATION NUMBER: US/10/084,846A  
CURRENT FILING DATE: 2003-02-25  
PRIOR APPLICATION NUMBER: PCT/EP01/09815  
PRIOR FILING DATE: 2001-08-24  
PRIOR APPLICATION NUMBER: DE 101 09 166.4  
PRIOR FILING DATE: 2001-02-25  
NUMBER OF SEQ ID NOS: 120  
SOFTWARE: PatentIn Ver. 3.2  
SEQ ID NO 4  
LENGTH: 19725  
TYPE: PRT  
ORGANISM: Streptomyces viridochromogenes  
FEATURE:  
OTHER INFORMATION: Protein 2: amino acid sequence encoded by coding strand 1.  
OTHER INFORMATION: Start codon: gat, Start position: nucleotide 2.  
US-10-084-846A-4

Alignment Scores:  
Pred. No.: 3,07e-16 Length: 19725  
Score: 407.00 Matches: 236  
Percent Similarity: 34.8% Conservative: 48  
Best Local Similarity: 29.0% Mismatches: 237  
Query Match: 11.2% Indels: 294  
Gaps: 53  
DB:

US-10-620-914-44 (1-1947) x US-10-084-846A-4 (1-19725)

QY	23	GGCTGGGAGCTACACCAAGAAAGAACTTCTCCCTGG-----AGAGCTCAAGC	70
DB	17261	GlyCysArgThrSerProArgAlaGlyCysAlaTrpGlyArgTrpArgAlaGlySer	17280
QY	71	TCAGCAGATGAAAGATGACCTGACCGTTCTGGCCCATATCTGGTTCGGCAGCAGAGG	130
DB	17281	GlyArgSer-----AlaArgCysSerThrProAlaArg---	17291
QY	131	CGGATGATCAGCTGCTGCTGGAGAGCTTCTAGGGCCCGCAGCCGCTGCTTGTCTG	190
DB	17292	-----ThrGlyCysAlaSerSerAlaGlySerGlySerArgPro-----	17304
QY	191	CCCGCC-----TGCCCGAGCGCTCGAACCTCATCTGGGTTGACCTGGTGGTGAC	244
DB	17305	ProThrArgArgTrpProProSerGlySerProArgLeu-----Tyr	17319
QY	245	GGGAGAATGTCCATATGATGCTGATTCATCGACCTGGCGAGTTCAAGTCCATCTACG	304
DB	17320	GlyThrAlaSer-----ProTyrThrSerThrTrpThrGlyProSerProGlyPro	17336
QY	305	-----TGCTCGACCTGTCGCTGTCGAGGTGCGAGGTGCGCAAGAAGA-----	346
DB	17337	ProHisTrpSer-----SerArgCysThrCysTrpTrpArgAlaProArgHis	17353
QY	347	-----AGCGAAGG-----CCAAAGGGCT	364
DB	17354	AlaGlyProSerArgAlaGlyArgArgSerSerAlaArgCysTrpProArgArg	17373
QY	365	GGAGA-----ATGTCAGTGC-----	382
DB	17374	GlyArgGlnAlaGlyCysAlaCysCysSerArgGlyGlyProHisGlyAspGlyArg	17393
QY	383	TGGAGCGCGACCTTGGCAATTTGGCCCTCGAGGCGACCG-----CGACGC	430
DB	17394	TrpArgProTrpArgSerAlaHisProSerArgProThrThrValCysHisArg	17413
QY	431	TCATCACCTTCTCTACTGCTCAGATGATTCACCGTTCCACACAGCTCATCGACGAG	490

DB	17632	ArgThrAlaSerArgSerAlaAlaTrpCys-----GlySerSerSer	17646
QY	889	TGCTCTTGGGGTTGATCTCCATCATCTCCATCCGGTCTGGGGTCTCCACAGCATGCG	830
DB	17647	ProSerTrp-----ArgArgProSerThrSer---	17655
QY	829	TGTACAGGAAGTGGCGGGNAACATGGCGGCCCTCCAGCGCTCTCTGTCGAGGCGGT	770
DB	17656	-----ThrThrAlaTrp-----ArgAlaProArg-----	17663
QY	769	GGCCCAAGCTGGGAGGCGGCATCCACAGTAGTAGGGGGCGC-----GCAGCC	719
DB	17664	-----ArgSerAlaArgSerHisArgSerIleCysAlaAla	17676
QY	718	ACGGCAGCTAGGCGATCGAACCTCGGTGTTCTCTCCACACGCGCTCCAGCTTCTGCT	659
DB	17677	ThrSerArgThrSerSerProAsnSerArgArgSerThrArgThrSerGlyArgSer	17696
QY	658	CCAGGTAGGCGCGCTCGGGCGCGATGTCATGTTGTCGATCGAAGATCATGCCGCC	599
DB	17697	ProAlaArgArg-----Thr	17701
QY	598	AGAAGAACGGCGCAGCAGGCGATCTGGCGAGGCGAGTCTGCTTCTCCAGCTCAGCT	539
DB	17702	ArgArgProThrAlaArgProCysAlaProAlaThrSerTrpArg-----CysAlaGlyArg	17720
QY	538	AGAACTCGGCAACGCCACACGCGCTCTGGGACAGGTACGACGACGCTGGTCGATGA	479
DB	17721	ArgThrArgValArgSerGlyProArgArgCysThrGlyCysAlaArg-----	17736
QY	478	CGTTGTGAACGGTGAATCATCTGCGAGCGATGAGAGAGCGTATGAGCGCTCGCGTGC	419
DB	17736	-----	17736
QY	418	CCTCAGGGGGCGCAATTGGCAAGCGTGGCTCCACGACCTTCTCCAGCCCT	359
DB	17737	-----SerLeuArgArgProProArgThrGlyArgSer-----	17747
QY	358	TGGCCTTCGCTCTCTCT-----TGCCCA-----CCTCGCACAGC	323
DB	17748	TrpCysSerProArgSerGlyThrTrpGlnTrpTrpAlaArgArgSerArgThrArg	17767
QY	322	AGTGCACAGCTGCACCACT-----AGATGGACTTGAACCTCGCAGGTGATGATCAG	266
DB	17768	ProGlyGlnGlyArgLysGlnAlaArgArgThrSerArgGlyProLysArg-----GlyGlu	17786
QY	265	CCATCATATCGACATCTCCCGAGTCCACCCAGCTCAACCCAGATGAGTTCCGAGC	206
DB	17787	ProGlnArgMetArgGlyAlaArg-----ArgTrpLysGlyGlnProArgArgThrArg	17805
QY	205	GCTCGG-----CCAGCGCGGCGACCAAGCGCAG-----CGG	176
DB	17806	AlaArgArgArgSerProProTyrGlyThrHisArgArgArgSerGlyArgGlyArg	17825
QY	175	CCTGGGGCGCGTAGAGCTCTCCAGCGAGCAGCGTATCATCGCCCTTCTGCTGCCGA	116
DB	17826	AlaGlyCysArgAspArgGlyProGlyValCysSerGlyArgSerArgAlaAlaCysArg	17845
QY	115	-----ACCACATATGGCGAGACGCTCAGGTATCTCTTCATCGC	77
DB	17846	ProProAlaAspSerSerSerThrThrSerProProArgProValThrProTyrCys	17865
QY	76	TGCTGAGCTTGAGCTTCTCCAGGGAGAGTCTCTTCTGTTGTTAGC-----	32
DB	17866	-----SerArgArgSerArgGlnAlaGluTrpAlaSerThrCysArgProPro	17881
QY	31	-----TCGAGCGCGCGCTCAGGACCGCCCA	2
DB	17882	ProTrpSerSerAlaSerArgSerSerArgPro	17893



; SEQ ID NO 3

; LENGTH: 19695

; TYPE: PRT

; ORGANISM: Streptomyces viridochromogenes

; FEATURE:

; OTHER INFORMATION: Protein 1: amino acid sequence encoded by coding strand 1.

; OTHER INFORMATION: Start codon: gga, Start position: nucleotide 1.

; US-10-084-846A-3

Alignment Scores:

Pred. No.: 8.56e-15 Length: 19695

Score: 383.50 Matches: 218

Percent Similarity: 34.2% Conservative: 57

Best Local Similarity: 27.1% Mismatches: 268

Query Match: 10.6% Indels: 260

DB: 4 Gaps: 48

US-10-620-914-44 (1-1947) x US-10-084-846A-3 (1-19695)

Qy 1945 AGTTGTCCT---TCTTGGCGC-----CCTTCGGGGGGCCATGTAGA 1907

Db 4930 SerCysValIleGlyTrpArgArgHisArgAlaAlaProProAlaCysLeuProCysAla 4949

Qy 1906 -----AGGAGCTGTACATGTTGACGGGTCCATGT-----AGCCTGAGTGG 1865

Db 4950 CysCysArgCysCysArgCysTrpProArgProCysCysGlyThrAlaProArgAla 4969

Qy 1864 CGCGCGGATGACAGCGCATGTCGAAGCCGCTTTCGATCAGCTCGCGGTAGGCGGGC 1805

Db 4970 Arg-----ThrThrArgArgAspProAlaProGlyValProArgArg-----Gly 4984

Qy 1804 TGAGGGAGCGGAGCGGAGATGAGTGCCTGCGCGCGGCGCACTGTGTCGCGGCACT 1745

Db 4985 MetArgArgMetSerArgArgGlyCysCysSerProSerProTrpTrp-----5001

Qy 1744 CGGCCAGCTGTGCGCACGGGCATATCCAGCCAGTCCACGTGTTCATCAGAA-----1691

Db 5002 -----TrpPro-----SerArgGlyProAlaGluArgTrp 5011

Qy 1690 -----TCACCTTGTGTAGTGCCTTTGAGTCTCTCCATGAAGAGTTGTGGAGA 1637

Db 5012 SerAlaGlyThrTrpAlaSer-----HisAlaTrpTrpAla 5023

Qy 1636 CGTTCAGTTGTCACACAGCCACTCTTGAGGTGGCGAAGCCGCTCGCGCAGGTAGG 1577

Db 5024 ArgSer-----ProGlySer 5028

Qy 1576 TGGGCGAGTTGTCGGCAGGAACCTTCCGGTGGAGGAGTGTGTAGTAGAAGTAGT-----1523

Db 5029 PheSerAlaLeuArg-----CysSerAlaSerSerLeuPro 5040

Qy 1522 TCTGCTGCGCAGTGCAGTGTCTCCGCCAGCGCTCCA-----1484

Db 5041 GlySerThrTrpProSerProProProSerThrProThrSerIleSerSerProArg 5060

Qy 1483 -----TGTGGCGCGATGATGTTCTCAATGGGATCGCGTCCGCTTGATCAGCGGT 1430

Db 5061 SerGlyTrpArgCysSerCysSerSerAlaTrpSerSerValThrProThrMetLysAla 5080

Qy 1429 ACTGCTTGGCGCGCAGCGCGCGCGAACCACGACGCGCTTGTGAAGAGCACCGGC 1370

Db 5081 ArgAlaAlaProGlyProProTrpTrpGluSerSerAlaSerAlaCys---ArgSerProSer 5099

Qy 1369 TCACGAACCTGACGAACA---GCCACACAGCGGCTTGGGCGGCTTCTTACGAAGTGA 1313

Db 5100 AlaValLeuTrpAspThrArgCysThrProAlaThrProArgThrAlaSerAlaSer---5118

Qy 1312 TCAGCATGTTGTCTCCACACAGCGGCTGCTCT-----CCATTGGGGGG 1265

Db 5119 -----CysArgSerProSerSerTrpAlaSerProAlaProProSerProCysTrpArg 5136

Qy 1264 CGTTGGCGAGGGCTTGACGGTCTTGCCAGTCTCCAGCACCACGCGGCGGCACTGCAGCA 1205

Db 5137 GlyCysTrpAsnAlaAlaGlyCysSerSerProGlyArgAlaProThrProSerTrpGlyPro 5156

Qy 1204 CCAGCACAGCTTGGCCATGCCCGCTGTGTAGTACAGCGCGTGTGGAAGTACACAGAGGC 1145

Db 5157 ProProProIleSerProAla-----GlyCysSerArgValSerSerProCysCys 5174

Qy 1144 GCTTGGACACAGAGTGTGTGCTGTGTTGCGACAGGAGCGCCAGCTTCTTCTCGTACA 1085

Db 5175 AlaAlaValArgProTrpAlaCysAlaProSerArgSerProArgSerSerAla 5194

Qy 1084 GCTCTCAATGCGCGGTGCACGCCCT-----1058

Db 5195 SerTrpTrpCysSerSerAlaProProCysAspAlaSerTrpSerGlyArgSerAlaGly 5214

Qy 1057 CGCGAACAGCTGCCACACGCTCTCAAACTCCAGTGT-----GAATGGCCACCT 1007

Db 5215 CysProThrAlaAlaCysSerHisSerSerProAlaSerCysCysProArgSerProPro 5234

Qy 1006 TCTTCAGCTCCAGAACGCGCGACTTGGCGGGTGTGACGTCCACCGACACCA-----956

Db 5235 SerSerAlaSerIleSerSerAlaProSerCysSerAlaProSerAlaArgArgPro 5254

Qy 955 -----CCTGGCGCGGCCCTCTGCACCA-----GCAGTTTCAGGCACTTCGAGC 914

Db 5255 ArgArgProTrpArgTrpProAlaAlaAsnCysArgAsnSerSerProProTrpCysCys 5274

Qy 913 CGCGCTAGTCA-----GGGTTCACGA 893

Db 5275 ArgArgSerSerProArgSerAlaArgProThrCysCysAsnSerAlaGlyAlaAla 5294

Qy 892 CCGTGCTCTTGGGTGTGATCTCCATCATCTCCATATCCGGTCCGGGTCTCCACAGACT 833

Db 5295 ProCysGlyCysGly-----ProAlaSer 5302

Qy 832 GCGTGTACAGAAAGTGGCGCGGAAATGGCGCGCGCTCCACCGCTCTCTCGTCGAGGG 773

Db 5303 ProCysTrpTrpProSerSerAlaAsnTrpArgAlaAlaArgArgGlnArgArgCys---5321

Qy 772 CGTGGCCACGCTGGCGAGCGGCCAATCCACAGTACGTAGTGGGGCGCGCAGCACCGCA 713

Db 5322 -----ArgTrpSerAla-----SerThrProCysAlaSerAlaCysSerThrAla 5336

Qy 712 -----CGTAGGGATCGAACCCCTGGGTGTCT 686

Db 5337 AlaAspProAsnTrpSerSerProSerGlySerAsnSerAlaSerHisProArgCysSer 5356

Qy 685 ---GCTCCACACCGCTCCAGCTTCTGCTCCAGTAGGCGCGCTCGGGCGCGAGT 629

Db 5357 ProCysTrpSerSerProCysAlaProProSerProArg-----ArgCys 5372

Qy 628 CAATGTTGT-----CGATGTCCAGATCGATCGCCAGCAAGAAACGCGCG 584

Db 5373 SerThrCysSerThrAlaProArgAlaProArgProProValGlnArgValPro 5392

Qy 583 ACCAGGCGCATCTGGCGCAGGCGCGAGTGTGCTGCTCCAGTACGTAGAGTTCGCAACGC 524

Db 5393 SerPheAlaAspAsp---GlyAlaAlaAsnSerAlaArgGluMetProGluArgValVal 5411

Qy 523 CCACGAGCGCTCTTGGGACA-----503

Db 5412 IleProSerArgIleProThrLysArgLeuSerSerLysGluCysArgSerValArgAla 5431

Qy 502 GGTACGACAGCCCTGGT-----CGATGACGTTGTGGAACGGTG 464

Db 5432 GlySerSerLysProGluAsnArgGluValGlyTrpCysArgProThrCysThrSerVal 5451

Qy 463 GAATCATCTGAGCGAGTAGGAGAGGTGATGAGCGTCCGCGTCCCTCAGGGGGCGCAA 404

Db 5452 LeuSerAlaSerSerTrpArgArgVal-----AlaSerAsnGlyProTrpAsnAla---5468

Qy 403 ATTGCAAGCGTGGCCTCCAGCACTTCCTTCAGCCCTTGGCTTCGCTTCT 344

Db 5469 ---AlaSerIleArgProLysArgProArgHisThrSerSerAlaGlySerProProAla 5487



QY 343 TCT-----TGCCCA----- 335  
 Db ThrSerProLeuArgLysTrpProCysGlyArgArgSerArgArgGlyGlyAlaAasp 5507  
 QY 334 -----CCTCGCACAGCGAGTGGCACA--- 314  
 Db ArgArgArgAenSerIleCysCysMetArgAlaProGlyThrArgAlaArgThrAla 5527  
 QY 313 GGTGACACAGTAGA-----TGGACTTGAACCT---TCGCCA----- 281  
 Db GlyArgProArgHisThrCysSerGlyThrTrpTrpAlaGlyThrCysSerProTrpArg 5547  
 QY 280 -----GGTCGATGTAATCAGCCATCATATTCGACATTCT 248  
 Db SerGlyArgAlaAlaThrAlaCysSerAlaAlaArgTrpAenSerProSer--AlaThrCysA 5567  
 QY 247 CCCAGTG-----CCACACCCAGGTCAACCCAGATGAGTTGCG 209  
 Db rgpProfileArgThrAlaProAlaArgSerProProThrThrThrGluArgArg-Ser 5586  
 QY 208 AGCGCTCGCCAGCGCGG-----CAGCAAAAGCGACGCGCGCTGGG 170  
 Db ThrGly-----GlyGlyTrpAlaProAlaSerSerAlaAlaMetArgSerArgProGly 5604  
 QY 169 GCCCGTAGA-----AGCTTCCAGGGGAGCAGCGTGATCATCCCTTCTTGTCTGCGGA 116  
 Db AlaAlaSerGlyProAlaSerProGlySerAlaArg---TrpProProArgGlyCysArg 5623  
 QY 115 ACCACATATGGCGAGACGGTCAAGTTCATCTTCATGCTGCTGAGTTGAGTTCTCCA 56  
 Db ArgSer-----SerArgCysThrAlaValThrSerArgCysSerArgAlaSerPro 5641  
 QY 55 GGG 53  
 Db 5642 Gly 5642  
 RESULT 14  
 US-10-118-495-3  
 ; Sequence 3, Application US/10118495  
 ; Publication No. US20030074688A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Benning, Christoph  
 ; APPLICANT: Riekhof, Wayne  
 ; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids  
 ; FILE REFERENCE: MSU-06897  
 ; CURRENT APPLICATION NUMBER: US/10/118,495  
 ; CURRENT FILING DATE: 2002-04-08  
 ; PRIOR APPLICATION NUMBER: 60/283,812  
 ; PRIOR FILING DATE: 2001-04-13  
 ; NUMBER OF SEQ ID NOS: 42  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 3  
 ; LENGTH: 416  
 ; TYPE: PRT  
 ; ORGANISM: Rhodobacter sphaeroides  
 US-10-118-495-3  
 Alignment Scores:  
 Pred. No.: 6.82e-15 Length: 416  
 Score: 379.50 Matches: 127  
 Percent Similarity: 43.1% Conservative: 68  
 Best Local Similarity: 28.1% Mismatches: 182  
 Query Match: 10.5% Indels: 75  
 DB: 4 Gaps: 20  
 US-10-620-914-44 (1-1947) x US-10-118-495-3 (1-416)  
 QY 691 ACCCAGGTTTCGATCCCTAGTCGCGTGGTGGCGGCCCTACTACTAGTGTGATGGC 750  
 Db 2 ThrGlnPheAlaLeuThrHisLeuPro-----AlaProPro-----ValAla 15

751 GCGTGTCCCGAGCGTTGGCCACGCCCTGCAC-----GAGAGCGC 789  
 Db Arg-----GlnIleGlyAlaAlaValHisArgThrSerLeuLeuSerAlaGlyLeu 33  
 790 GTGAGAGCGCGCCCATGTTCCCGCCACCTTC-----CTGTACACGAGTCGTGG 840  
 QY 34 MetGluArg-----MetPheSerArgLeuPheHisGlyLeuValTrpProGlnIleTrp 51  
 Db 841 GAGGACCCCGAGCGGATATGGAGGTGATGAGATCAACCCCAAGGACACAGTGTGTGACC 900  
 QY 52 GluAaspProAlaValAaspMetAlaAlaLeuAlaIleArgProGlyAaspArgLeuValAla 71  
 Db 901 CTGACTAGCGCGCTGCAATGCCCTGAACCTGTGTGTCAGGGGCGCGCCAGGTGTGTG 960  
 QY 72 IleAlaSerGlyGlyCysAenValLeuSerTrpLeuThrGlnGlyProGlySerIleLeu 91  
 Db 961 TCGGTGAGCTGCACCCCGCGCAGTCGCGCTTCTGGAGCTGAAGAAGGTGGCGCATTCAG 1020  
 QY 92 AlaValAaspLeuSerProAlaHisValAlaLeuGlyArgLeuHisLeuAlaAlaArg 111  
 Db 1021 CAGCTG---GAGTTTGGAGCAGTGTGGCAGCTGTTCCGCGAGGCGGTGCACCCCGCAT 1077  
 QY 112 ThrLeuProAaspHisAlaAlaPhePheAaspLeuPheGlyArgAlaAaspLeuProGlyAen 131  
 Db 1078 GAGGAGCTGTACGAGAGAGAGTGGCGCCCTTCTGTGCGCAACACGACCACTCTTGG 1137  
 QY 132 AlaAlaLeuTrpAaspArgHisIleAlaProAlaLeuAaspGlyArgSerArgTrp 151  
 Db 1138 -----TCCAGCGCTCTGGTACTTCCAGCACGCGCTGTACTACCAG 1179  
 QY 152 GluAlaArgSerProPheGlyArgArgIleGlnLeuPheGluArgGlyPheTrpArgHis 171  
 Db 1180 GCGCGCATGGCAAGCTGTGTGGTGTGCAGTGCCTGCGCGGTGGTGTGGAGTGGCG 1239  
 QY 172 GlyAlaLeuGlyArgPheIleGlyAlaAlaHisThrLeuAla-----ArgAlaAaspGly 189  
 Db 1240 AAGACCTCAAGCGCTTCGCCAACCGCCCAATAGGAGAGAGCGCGCTGTGTGGAC 1299  
 QY 190 ThrAaspLeuArgGlyPheLeuAaspCysProAaspIleGluAlaGlnArgSerPheTrp 209  
 Db 1300 AGCAACATGTCTCCTCCACTTCGTGAAGAAGCGGCGCCAGCGCTGTGTGTGTGTGTGTC 1359  
 QY 210 AlaHisIle-----GlyPro-----LeuPheGlu 217  
 Db 1360 AAGTTCGTGAGCGCTGTGCTCTTCAACAAGCGCGCTGTGTGTGTGTGTGTGTGTGTGTG 1419  
 QY 218 AlaProValValGlnAlaLeuAlaArgProAlaAlaLeuPheGlyLeuGlyLeuPro 237  
 Db 1420 GGCAGCAGTACGCGCTGATCAAGCGGACGCG-----ATCCCATTTGAGAAC 1467  
 QY 238 ProAlaGlnTrpAlaLeuLeuAlaGlyAaspGlyAaspValLeuProVal----- 255  
 Db 1468 TACATCGCGCGCACCATGGAGCGGTGGCGGAGACTCGCAC----- 1509  
 QY 256 -----LeuArgGlnArgLeuHisArgLeuLeuCysAaspPhe 267  
 Db 1510 GTGCGCAAGCAGAACTACTTCTACTCAACTGCTCACCAGCAAGTTCTCTGCGCGCAAC 1569  
 QY 268 ProLeuArgGluAenTrpPheAlaPheGlnAlaIleAlaArgArgTrpProArgProGly 287  
 Db 1570 -----TGCCCGACCTTACCTGCGGAGCGCGCTTCCACCTCCAGAGTGGCGGTG 1620  
 QY 288 GluGlyAlaLeuProTrpTrpAlaPheGluThrLeuArgGluAen--- 306  
 Db 1621 GTGACAACTGACCGCTCTCCCACTTCTCATGAGGAGGTCAAAGCGCGCACCTAC 1680  
 QY 307 AlaGlyArgValGlnIleGluAenArgSerLeuThrGluAlaLeuAlaGluProGlu 326  
 Db 1681 ACCAAGGTG-----ATTCTGATGACACAGCTGAGCTGGCTGATATGCCGTGGCC 1731  
 QY 327 GluSerIleHisGlyPheThrLeuLeuAaspAlaGlnAaspTrpMetThrAaspAlaGlnLeu 346  
 Db 1732 AACGAGCTGCGCGAGTGCCTGGCCAAAGCAGAGTGTGGCGCGCGCGCATCTGTCATCTGCGCGC 1791

Db 347 ThrAlaLeuTrpArgGlnValThrArgThrAlaAlaProGlyAlaArgValIlePheArg 366  
Qy 1792 -----TCGCGCTCCCTCAGCCCGCCCTACCGCGAGCTGATCCAGAGCGGGCTTC 1842  
Db 367 ThrGlyGlyAlaAlaAspLeuPro-----GlyArgValProGluGluLeuGly 384  
Qy 1843 GACGTGGCTGCATCCCGCGCCCACTCAGGCG-----TACATGGACCGGTCAACATG 1896  
Db 385 HisTrpArgAlaAspArgAlaAlaGlyGlnAlaGlyHisAlaAlaAspArgSerAlaIle 404  
Qy 1897 TACAGCTCTTCTACATGGCCCGCGGAGGCGCC 1932  
Db 405 TyrGlyPheHisLeuTyrArgArgAspAla 416

RESULT 15  
US-10-620-914-3  
; Sequence 3, Application US/10620914  
; Publication No. US20040093639A1  
; GENERAL INFORMATION:  
; APPLICANT: Benning, Christoph  
; APPLICANT: Riekhof, Wayne  
; APPLICANT: Klug, Rouven  
; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids  
; FILE REFERENCE: MSU-07769  
; CURRENT APPLICATION NUMBER: US/10/620,914  
; CURRENT FILING DATE: 2003-07-16  
; PRIOR APPLICATION NUMBER: 10/118,495  
; PRIOR FILING DATE: 2002-04-08  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 3  
; LENGTH: 416  
; TYPE: PRT  
; ORGANISM: Rhodobacter sphaeroides  
US-10-620-914-3

Alignment Scores:  
Pred. No.: 6.82e-15 Length: 416  
Score: 379.50 Matches: 127  
Percent Similarity: 43.1% Conservative: 68  
Best Local Similarity: 28.1% Mismatches: 182  
Query Match: 10.5% Indels: 75  
Gaps: 20

US-10-620-914-44 (1-1947) x US-10-620-914-3 (1-416)

Qy 691 ACCGAGGTTCGATCCCTACGTGCGGTGGCTGCGCGCCCTACTACGTGTGGATTGGC 750  
Db 2 ThrGlnPheAlaLeuThrHisLeuPro-----AlaProPro-----ValAla 15  
Qy 751 CGCTGCGCAGGTTGGCCAGCCCTGCAC-----GAGGAGCGC 789  
Db 16 Arg-----GlnIleGlyAlaAlaValHisArgThrSerLeuLeuSerAlaGluGlyLeu 33  
Qy 790 GTGGAGCGCGCCCATGTTCCCGCCCACTTC-----CTGTACAGCGAGTCGTGG 840  
Db 34 MetGluArg-----MetPheSerArgLeuPheHisGlyLeuValTyrProGlnIleTrp 51  
Qy 841 GAGGACCCCGAGCGCGATGATGAGGTGATGGAGATCAACCCCAAGGACACCGGTGTGACC 900  
Db 52 GluAspProAlaValAspMetAlaAlaLeuAlaIleArgProGlyAspArgLeuValAla 71  
Qy 901 CTGACTAGCGCGGTGCAATGCCCTGAACTGTGGTGCAGGGGCGCGCGAGTGTGG 960  
Db 72 IleAlaSerGlyGlyCysAsnValLeuSerTyrLeuThrGlnGlyProGlySerIleLeu 91  
Qy 961 TCGTGGAGTGAACCCCGCGAGTCGCGGCTTCTGGAGCTCAAGAGTGGCCATTGAG 1020  
Db 92 AlaValAspLeuSerProAlaHisValAlaLeuGlyArgLeuLeuLeuAlaAlaArg 111  
Qy 1021 CAGCTG---GAGTTTGGAGCAGTGTGGCAGCTGTTCGGCGAGGCGGTGCACCCGCGCAT 1077

Db 112 ThrLeuProAspHisAlaAlaPhePheAspLeuPheGlyArgAlaAspLeuProGlyAsn 131  
Qy 1078 GAGGAGTGTACGAGAAGAGTGGCGCCCTCTCTGTCGCAACACGACCACTTCTGG 1137  
Db 132 AlaAlaLeuTyrAspArgHisIleAlaProAlaLeuAspGlyArgSerArgTyrTrp 151  
Qy 1138 -----TCCAAAGCGCTCTGGTACTTCCAGCACGCGCCTGTACTACCAG 1179  
Db 152 GluAlaArgSerProPheGlyArgGileGlnLeuPheGluArgGlyPheTyrArgHis 171  
Qy 1180 GCGCGCATGGCAAGTGTCTGGTCTGCTGAGTGCCTGCGCGGTGGTGGAGTGGCG 1239  
Db 172 GlyAlaLeuGlyArgPheIleGlyAlaAlaHisThrLeuAla-----ArgAlaAlaGly 189  
Qy 1240 AAGACCGTCAAGCGCTCGCCACCGCCACATGAGGAGGAGCGCGCTCTGGGAC 1299  
Db 190 ThrAspLeuArgGlyPheLeuAspCysProAspIleGluAlaGlnArgSerPhePheTyr 209  
Qy 1300 AGCAACATGCTCATCCACTTCTGTAAGAAGCGGCGCCAAAGCGCTGTGGTGTGCTC 1359  
Db 210 AlaHisIle-----GlyPro-----LeuPheGlu 217  
Qy 1360 AAGTTCGTGAGCCGTGTCTTCAACAGCCCGTGTGTGTCGCGCGCGGTGGCG 1419  
Db 218 AlaProValValGlnAlaLeuAlaArgArgProAlaAlaLeuPheGlyLeuGlyIlePro 237  
Qy 1420 GGCAGCAGTACGCGGTGATCAAGCGCGAGCGC-----ATCCCATTTGAGAAC 1467  
Db 238 ProAlaGlnTyrAlaLeuLeuAlaGlyAspGlyAspValLeuProVal----- 255  
Qy 1468 TACATCGCGCGCACCATGAGCGGTGGCGGAGCAACTCGCAC----- 1509  
Db 256 -----LeuArgGlnArgLeuHisArgLeuLeuCysAspPhe 267  
Qy 1510 GTGCGCAAGCAGAACTACTTACTACAACTGCTCACCAGCAAGTTCCTCGCGGACAC 1569  
Db 268 ProLeuArgGluAsnTyrPheAlaPheGlnAlaIleAlaArgArgTyrProArgProGly 287  
Qy 1570 -----TGCCCGCCTACCTCGCGGAGCGCGCTTCCGACCCCTCAAGAGTGGCGTG 1620  
Db 288 GluGlyAlaLeuProProTyrLeuGluProThrAlaPheGluThrLeuArgGluAsn--- 306  
Qy 1621 GTGCAACACTTGACCGCTCCACCAACTTCTTCATGAGGAGAGCTCAAGCGCGCACCTAC 1680  
Db 307 AlaGlyArgValGlnIleGluAsnArgSerLeuThrGluAlaLeuAlaAlaGluProGlu 326  
Qy 1681 ACCAAGGTG-----ATTCTGATGACACAGTGGAGTGTGCTGGATATGCCGTGGCC 1731  
Db 327 GluSerIleHisGlyPheThrLeuLeuAspAlaGlnAspTrpMetThrAspAlaGlnLeu 346  
Qy 1732 AACGAGTGGCGGAGTGTGCGCAAGCAGGTTCGCGCGCGGCGCATCGTCTATCTGGCGC 1791  
Db 347 ThrAlaLeuTrpArgGlnValThrArgThrAlaAlaProGlyAlaArgValIlePheArg 366  
Qy 1792 -----TCGCGCTCTCTCAGCGCCCTACGCGGAGCTGATCCAGAGAGCGGCTTC 1842  
Db 367 ThrGlyGlyAlaAlaAspLeuPro-----GlyArgValProGluGluLeuGly 384  
Qy 1843 GACGTGCTGTGATCCCGCGCCACTCAGGCG-----TACATGAGCGCGTCAACATG 1896  
Db 385 HisTrpArgAlaAspArgAlaAlaGlyGlnAlaGlyHisAlaAlaAspArgSerAlaIle 404  
Qy 1897 TACAGCTCTTCTACATGGCCCGCGGAGGCGCC 1932  
Db 405 TyrGlyPheHisLeuTyrArgArgAspAla 416

Search completed: March 14, 2006, 01:57:05  
Job time : 459.5 secs

Result No.	Score	Query Match	Length	DB	ID	Description
1	3463	100.0	648	4	US-10-620-914-45	Sequence 45, Appl
2	891	25.7	752	5	US-10-741-849-7021	Sequence 7021, Ap
3	850.5	24.6	908	4	US-10-620-914-50	Sequence 50, Appl
4	379.5	11.0	416	4	US-10-118-495-3	Sequence 3, Appli
5	379.5	11.0	416	4	US-10-620-914-3	Sequence 3, Appli
6	326.5	9.4	416	4	US-10-118-495-29	Sequence 29, Appl
7	326.5	9.4	416	4	US-10-620-914-29	Sequence 29, Appl
8	321	9.3	415	4	US-10-118-495-41	Sequence 41, Appl
9	321	9.3	415	4	US-10-620-914-41	Sequence 41, Appl
10	292.5	8.4	416	4	US-10-118-495-33	Sequence 33, Appl
11	292.5	8.4	416	4	US-10-620-914-33	Sequence 33, Appl
12	123	3.6	2284	5	US-10-684-141-58	Sequence 58, Appl
13	123	3.6	2284	5	US-10-810-486-58	Sequence 58, Appl
14	120	3.5	221	4	US-10-118-495-35	Sequence 35, Appl
15	120	3.5	221	4	US-10-620-914-35	Sequence 35, Appl
16	119	3.4	2286	5	US-10-684-141-54	Sequence 54, Appl
17	119	3.4	2286	5	US-10-810-486-54	Sequence 54, Appl
18	112.5	3.2	672	4	US-10-282-122A-68178	Sequence 68178, A
19	111	3.2	198	5	US-10-617-320-4622	Sequence 4622, Ap
20	110.5	3.2	298	5	US-10-450-763-38334	Sequence 38334, A
21	107	3.1	255	4	US-10-220-381-24	Sequence 24, Appli
22	107	3.1	255	5	US-10-988-984-2	Sequence 2, Appli
23	107	3.1	691	5	US-10-732-923-9665	Sequence 9665, Ap
24	106.5	3.1	670	4	US-10-282-122A-69771	Sequence 69771, A
25	106	3.1	472	3	US-09-974-298-120	Sequence 120, App
26	106	3.1	472	4	US-10-247-671-186	Sequence 186, App
27	106	3.1	472	4	US-10-211-462-201	Sequence 201, App



Db 478 HWRIYDDQLPKTQFNDEYIYAFTWEDSRVRELLNLGPDVVLAITSAGDNLSYLMQ 537  
Qy 315 GAGQVVSVDNCPAQSALELKKVAIOLEPEDVWQFGEVHPRIEBELYEKKLAPFLSQT 374  
Db 538 SPARVAIDLNPQNHLELKVASFITLDPDVWVKIFGEGKHPDFRSLLSKLSPHLSGR 597  
Qy 375 SHNFWKRLWYFO-----HGLYOGGKGLCWVLCQAVLGLGKTVKRLANAPTMEQRR 430  
Db 598 AFQYWLNSAHIFTDPAGRLYDGGSRYAIRFRWISTLFFCRSAVRRLSTPTEGQRS 657  
Qy 431 LWDNMLIHVPKNGPPLVWLFVKFYS-LVFNKAVLWFGGVPKQYALIKAD-----483  
Db 658 IYHTKI-----RPC-LIMRPNGLVSSDAFLMSALGPKQVQVAMIEADYHRSI 706  
Qy 484 -----GPIENIYARTMDGVAENSHVRKQNYFYNNCLTGKFLRNCPTYL 529  
Db 707 SSTTSSKEKPSRAEAILHYTSTLDPVLSLSDNPYLVCLVQVYTRQCHDYLS 766  
Qy 530 EAPATLKS-GVVDNLTVSTNFMEL-----KARTYKVLIMDHWDMW 575  
Db 767 PAAHSILSAPGAFDGLRIHTDEIQELARFQGLTVAVVMDGMDWFPSPPEEKEGRG 826  
Qy 576 VANELAECLAQVAPGIVWRSASLSPPYAELIQAGPDVRCI-----RRATQGYMDRV 630  
Db 827 KAREQVRRRLNRAUKVGKVLRSAGVEPTVYRVFVBEGFGARRVCGRESGRGQBCIDRV 886  
Qy 631 NMYSSFYMAR 641  
Db 887 NMYASCMILEK 897

## RESULT 4

US-10-118-495-3  
; Sequence 3, Application US/10118495  
; Publication No. US20030074688A1  
; GENERAL INFORMATION:  
; APPLICANT: Benning, Christoph  
; APPLICANT: Riekhof, Wayne  
; APPLICANT: Klug, Rouven  
; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids  
; FILE REFERENCE: MSU-06897  
; CURRENT APPLICATION NUMBER: US/10/118,495  
; CURRENT FILING DATE: 2002-04-08  
; PRIOR FILING DATE: 60/283,812  
; PRIOR FILING DATE: 2001-04-13  
; NUMBER OF SEQ ID NOS: 42  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 3  
; LENGTH: 416  
; TYPE: PRT  
; ORGANISM: Rhodobacter sphaeroides  
US-10-118-495-3

Query Match 11.0%; Score 379.5; DB 4; Length 416;  
Best Local Similarity 28.1%; Pred. No. 8.3e-29;  
Matches 127; Conservative 68; Mismatches 182; Indels 75; Gaps 20;

Qy 231 TQGSIPVPLRPYVWIGRLPSVGHALH-----EERVPMPFPPTF---LYTQSW 280  
Db 2 TPALHLP---APP---VAR--QIGAAVHRTSLLSAEGLMER--MFSRLFHGLVYVQIW 51  
Qy 281 EDPEPDMVMEINPKDTVLTLTSGGCNVALNLVQAGQVVSVDNCPAQSALELKKVAIQ 340  
Db 52 EDPAVDMALAIRPGDLVAIASGGCNVLSYLTQGGSTLAVDLSPAHVALGRLKLAAR 111  
Qy 341 QL-EFEDVWOLFGEVHPRIEELYEKKLAPFLSQTSHNF-----SKRLWYFQHGLYQ 393  
Db 112 TLPDHAFFDLGRADLPNGAALYDRHIAPALDGRSRRRYWEARSPPGRRIQLFERFYRH 171  
Qy 394 GGMKLCWVLCQAVLGLGKTVKRLANAPTMEEQRLWDSNMLIHFKVNGPKPLVWLFV 453  
Db 172 GALGRFIGAAHTLA---RAAGTDLRGFLDCPDIEAQRFFVYHI-----GP-----LPE 217  
Qy 454 KFSVSLVLFNKAVLWFGGVPKQYALIKADG-----IPIENIYARTMDGVAENSH-----503  
Db 218 APVQALARRPAALFGLGIPPAQYALLAGDGDVLPV-----LRQRLHLLCDF 267  
Qy 504 VRKQNYFYNNCLTGKFLRDN---CPTVLRFAAFATLKSQVVDNLTVSTNFFMELKARTY 560  
Db 268 PIRENTFAQAIARPRPRGEGALPYLEPTAFETUREN-AGRVQIENSLTALAAEPE 326  
Qy 561 TKV---ILMDHVDWLMVPVANELAECLAQVAPGGIVWIR---SASLSPPYAELIQAGF 614  
Db 327 ESHGFTLLDAQDWMTDALQTLWRQVTRTAAPGARVIFRTGGAADLLP--GRVPBEILG 384  
Qy 615 DVRCIRRAQOG--YMDRVNMYSSFYMAR 644  
Db 172 GALGRFIGAAHTLA---RAAGTDLRGFLDCPDIEAQRFFVYHI-----GP-----LPE 217

Qy 454 KFSVSLVLFNKAVLWFGGVPKQYALIKADG-----IPIENIYARTMDGVAENSH-----503  
Db 218 APVQALARRPAALFGLGIPPAQYALLAGDGDVLPV-----LRQRLHLLCDF 267  
Qy 504 VRKQNYFYNNCLTGKFLRDN---CPTVLRFAAFATLKSQVVDNLTVSTNFFMELKARTY 560  
Db 268 PIRENTFAQAIARPRPRGEGALPYLEPTAFETUREN-AGRVQIENSLTALAAEPE 326  
Qy 561 TKV---ILMDHVDWLMVPVANELAECLAQVAPGGIVWIR---SASLSPPYAELIQAGF 614  
Db 327 ESHGFTLLDAQDWMTDALQTLWRQVTRTAAPGARVIFRTGGAADLLP--GRVPBEILG 384  
Qy 615 DVRCIRRAQOG--YMDRVNMYSSFYMAR 644  
Db 385 HWRADRAAGQAGHAADRSAYIGGFHLYRRDA 416

## RESULT 5

US-10-620-914-3  
; Sequence 3, Application US/10620914  
; Publication No. US20040093639A1  
; GENERAL INFORMATION:  
; APPLICANT: Benning, Christoph  
; APPLICANT: Riekhof, Wayne  
; APPLICANT: Klug, Rouven  
; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids  
; FILE REFERENCE: MSU-07769  
; CURRENT APPLICATION NUMBER: US/10/620,914  
; CURRENT FILING DATE: 2003-07-16  
; PRIOR FILING DATE: 2003-07-16  
; PRIOR FILING DATE: 2002-04-08  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 3  
; LENGTH: 416  
; TYPE: PRT  
; ORGANISM: Rhodobacter sphaeroides  
US-10-620-914-3

Query Match 11.0%; Score 379.5; DB 4; Length 416;  
Best Local Similarity 28.1%; Pred. No. 8.3e-29;  
Matches 127; Conservative 68; Mismatches 182; Indels 75; Gaps 20;

Qy 231 TQGSIPVPLRPYVWIGRLPSVGHALH-----EERVPMPFPPTF---LYTQSW 280  
Db 2 TPALHLP---APP---VAR--QIGAAVHRTSLLSAEGLMER--MFSRLFHGLVYVQIW 51  
Qy 281 EDPEPDMVMEINPKDTVLTLTSGGCNVALNLVQAGQVVSVDNCPAQSALELKKVAIQ 340  
Db 52 EDPAVDMALAIRPGDLVAIASGGCNVLSYLTQGGSTLAVDLSPAHVALGRLKLAAR 111  
Qy 341 QL-EFEDVWOLFGEVHPRIEELYEKKLAPFLSQTSHNF-----SKRLWYFQHGLYQ 393  
Db 112 TLPDHAFFDLGRADLPNGAALYDRHIAPALDGRSRRRYWEARSPPGRRIQLFERFYRH 171  
Qy 394 GGMKLCWVLCQAVLGLGKTVKRLANAPTMEEQRLWDSNMLIHFKVNGPKPLVWLFV 453  
Db 172 GALGRFIGAAHTLA---RAAGTDLRGFLDCPDIEAQRFFVYHI-----GP-----LPE 217  
Qy 454 KFSVSLVLFNKAVLWFGGVPKQYALIKADG-----IPIENIYARTMDGVAENSH-----503  
Db 218 APVQALARRPAALFGLGIPPAQYALLAGDGDVLPV-----LRQRLHLLCDF 267  
Qy 504 VRKQNYFYNNCLTGKFLRDN---CPTVLRFAAFATLKSQVVDNLTVSTNFFMELKARTY 560  
Db 268 PIRENTFAQAIARPRPRGEGALPYLEPTAFETUREN-AGRVQIENSLTALAAEPE 326  
Qy 561 TKV---ILMDHVDWLMVPVANELAECLAQVAPGGIVWIR---SASLSPPYAELIQAGF 614  
Db 327 ESHGFTLLDAQDWMTDALQTLWRQVTRTAAPGARVIFRTGGAADLLP--GRVPBEILG 384  
Qy 615 DVRCIRRAQOG--YMDRVNMYSSFYMAR 644  
Db 172 GALGRFIGAAHTLA---RAAGTDLRGFLDCPDIEAQRFFVYHI-----GP-----LPE 217

Db 385 HWRADRAAGQAHAADRSAIYGGFHLRYRRDA 416

RESULT 6

US-10-118-495-29

Sequence 29, Application US/10118495

Publication No. US20030074688A1

GENERAL INFORMATION:

APPLICANT: Benning, Christoph

APPLICANT: Riekhof, Wayne

APPLICANT: Klug, Rouven

TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids

FILE REFERENCE: MSU-06897

CURRENT APPLICATION NUMBER: US/10/118,495

CURRENT FILING DATE: 2002-04-08

PRIOR APPLICATION NUMBER: 60/283,812

PRIOR FILING DATE: 2001-04-13

NUMBER OF SEQ ID NOS: 42

SOFTWARE: PatentIn version 3.1

SEQ ID NO 29

LENGTH: 416

TYPE: PRT

ORGANISM: Agrobacterium tumefaciens

US-10-118-495-29

Query Match 9.4%; Score 326.5; DB 4; Length 416;

Best Local Similarity 27.8%; Pred. No. 1.9e-23;

Matches 111; Conservative 69; Mismatches 158; Indels 61; Gaps 16;

Qy 275 LYTSQWEDPEPMEVMEINPKDVTLTSTGGCNALNLLVQAGQGVVSDCNPAQSALLEL 334

Db 44 VYPQIWEDEPIDMEAMELGEGHRIVTIGSGCNMLAYLSRNPASIDVVDLNPPIHIALNKL 103

Qy 335 KKVAIQOL-EFEDVMQLFCEGVHPRIEELYEKLAAPFLSQTSHNFWSK-----RLWYF 386

Db 104 KLAARFHLPAHQDVVRHFRAGTRNSVGYDRFAIEHLDAITKAYSKRTLSGRRRISVF 163

Qy 387 OHGLYYQGGMGKLCWVLOCLAVVLGLGKTVKRLANAPTMEEQRRLWDSNMLIHFKVNGPK 446

Db 164 DRNIYRTGLGRFIGAGHIMARLHGKLT--EMAKTRTLDQORQFDSKVAPLP----DK 217

Qy 447 PLV-WLFFVKFVSILVFNKAVLWFGGVPKQY---ALIKADGIPENIYARTMDGVAENS 502

Db 218 PVVRWLTGRKSSL-----FGLGIPRQYDELASLSSDG-TVASVLKERLEKACNF 267

Qy 503 HVRKQNYFYNNCLTGKFLRDN---CPTYLREAAATLKSQVVDNLTVSTNFFMEELK--- 556

Db 268 PL-SDNYFAWQAPARRYPPEHGAIPAYLKPEYKIRNNTA-RVAVHATYTELLSRKP 325

Qy 557 ARTYTKVILMDHVDLMDPMEVANELAECLAKQVAPGGIVWRSAS-----LSPPYAEL 608

Db 326 ANGVDRIYLLDAQDWMVDVQNELMSQISRTAASGARVIFRTAAEKSVIEGRSLP----- 380

Qy 609 IQKAGFDVR-----CIRRATQ-GYMDRVNMYSSFFMARR 641

Db 381 -----DIRNQWVYLEERSNELNAMDRAISYGGFHIYOR 413

RESULT 7

US-10-118-495-29

Sequence 29, Application US/10620914

Publication No. US20040093639A1

GENERAL INFORMATION:

APPLICANT: Benning, Christoph

APPLICANT: Riekhof, Wayne

APPLICANT: Klug, Rouven

TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids

FILE REFERENCE: MSU-07769

CURRENT APPLICATION NUMBER: US/10/620,914

CURRENT FILING DATE: 2003-07-16

PRIOR APPLICATION NUMBER: 10/118,495

PRIOR FILING DATE: 2002-04-08

NUMBER OF SEQ ID NOS: 52

SOFTWARE: PatentIn version 3.1

SEQ ID NO 29

LENGTH: 416

TYPE: PRT

ORGANISM: Agrobacterium tumefaciens

US-10-118-495-29

Query Match 9.4%; Score 326.5; DB 4; Length 416;

Best Local Similarity 27.8%; Pred. No. 1.9e-23;

Matches 111; Conservative 69; Mismatches 158; Indels 61; Gaps 16;

Qy 275 LYTSQWEDPEPMEVMEINPKDVTLTSTGGCNALNLLVQAGQGVVSDCNPAQSALLEL 334

Db 44 VYPQIWEDEPIDMEAMELGEGHRIVTIGSGCNMLAYLSRNPASIDVVDLNPPIHIALNKL 103

Qy 335 KKVAIQOL-EFEDVMQLFCEGVHPRIEELYEKLAAPFLSQTSHNFWSK-----RLWYF 386

Db 104 KLAARFHLPAHQDVVRHFRAGTRNSVGYDRFAIEHLDAITKAYSKRTLSGRRRISVF 163

Qy 387 OHGLYYQGGMGKLCWVLOCLAVVLGLGKTVKRLANAPTMEEQRRLWDSNMLIHFKVNGPK 446

Db 164 DRNIYRTGLGRFIGAGHIMARLHGKLT--EMAKTRTLDQORQFDSKVAPLP----DK 217

Qy 447 PLV-WLFFVKFVSILVFNKAVLWFGGVPKQY---ALIKADGIPENIYARTMDGVAENS 502

Db 218 PVVRWLTGRKSSL-----FGLGIPRQYDELASLSSDG-TVASVLKERLEKACNF 267

Qy 503 HVRKQNYFYNNCLTGKFLRDN---CPTYLREAAATLKSQVVDNLTVSTNFFMEELK--- 556

Db 268 PL-SDNYFAWQAPARRYPPEHGAIPAYLKPEYKIRNNTA-RVAVHATYTELLSRKP 325

Qy 557 ARTYTKVILMDHVDLMDPMEVANELAECLAKQVAPGGIVWRSAS-----LSPPYAEL 608

Db 326 ANGVDRIYLLDAQDWMVDVQNELMSQISRTAASGARVIFRTAAEKSVIEGRSLP----- 380

Qy 609 IQKAGFDVR-----CIRRATQ-GYMDRVNMYSSFFMARR 641

Db 381 -----DIRNQWVYLEERSNELNAMDRAISYGGFHIYOR 413

RESULT 8

US-10-118-495-41

Sequence 41, Application US/10118495

Publication No. US20030074688A1

GENERAL INFORMATION:

APPLICANT: Benning, Christoph

APPLICANT: Riekhof, Wayne

APPLICANT: Klug, Rouven

TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids

FILE REFERENCE: MSU-06897

CURRENT APPLICATION NUMBER: US/10/118,495

CURRENT FILING DATE: 2002-04-08

PRIOR APPLICATION NUMBER: 60/283,812

PRIOR FILING DATE: 2001-04-13

NUMBER OF SEQ ID NOS: 42

SOFTWARE: PatentIn version 3.1

SEQ ID NO 41

LENGTH: 415

TYPE: PRT

ORGANISM: Mesorhizobium loti

US-10-118-495-41

Query Match 9.3%; Score 321; DB 4; Length 415;

Best Local Similarity 26.2%; Pred. No. 6.7e-23;

Matches 113; Conservative 70; Mismatches 176; Indels 72; Gaps 18;

Qy 255 VGHALHEERV-----ERPPMPFPPTF---LYTSQWEDPEPMEVMEINPKDVTLTSTG 304

Db 16 VGVAVQNRALSQAGISER--LFAFLFSLVYVQIWEDEPIDMEAMELGEGHRIVTIASG 73

Qy 305 GCNALLVQAGQGVVSDCNPAQSALLEKKAIVQOLEFE-DVMQLFGEVHPRIEEL 363

Db 74 GCNLLAYLRSAPRIDAVDLNAAHIALNKLKLEAVRLPFSQGLDFRFFGAADTSHNSQAY 133



QY 364 EKKLAPFLSQTSHNFWKRLW-----YFOHGLYYQGGMGKLCWLQCLAVVLG-----411  
DB 134 DRFIAPHLDPVSRHYWERNRGRRIAVFDRNFYQTLGLG-----LFTAMGHRTAK 185  
QY 412 -LGKTVKRLANAPTMEQORLWDSNMLIHFKVNGPKPLWLFVKFVSLVLFNKAVLWFGG 470  
DB 186 FFGVNPAPHMWEARNIGEQRRFNEELAPVFDK---KLLKWATSRKASL-----FGL 233  
QY 471 GVPKQY--ALIKADGIPIENTYAR-----TMDGVAENSHVRKQNYFYNYNCLTGKFLRDN- 523  
DB 234 GIPPAQYDSLITSGDGTMSVLKARLEKACDFPLEN-----NYFAWQAFARRYNPGE 287  
QY 524 --CPTYLREAAAFATLKSQGVVDNLTVSTNFFMBEL---KARTYTKVILMDHVDWLDMPVAN 578  
DB 288 AALPAYLEKQNTETIR-GNIDRVAIHANLIEFLAGKADGTVDRLFLLDAQDMTDDQLN 346  
QY 579 ELAECLAKQVAPGGIWIWSA---SLSP--PYAELIQKAGFVRCIRRTATQGYMDRVNMY 633  
DB 347 ALWSEISRSTASAGARVIFRTAAEPSLLPGRVSTSLDDQWDYQDEASREFSA--RDRSAIY 404  
QY 634 SSFYMARRKGA 644  
DB 405 GGFHLYVVKETA 415

## RESULT 9

US-10-620-914-41  
; Sequence 41, Application US/10620914  
; Publication No. US20040093639A1  
; GENERAL INFORMATION:  
; APPLICANT: Benning, Christoph  
; APPLICANT: Riekhof, Wayne  
; APPLICANT: Klug, Rouven  
; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids  
; FILE REFERENCE: MSU-07769  
; CURRENT APPLICATION NUMBER: US/10/620,914  
; CURRENT FILING DATE: 2003-07-16  
; PRIOR FILING DATE: 2002-04-08  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 41  
; LENGTH: 415  
; TYPE: PRT  
; ORGANISM: Mesorhizobium loti  
US-10-620-914-41

Query Match 9.3%; Score 321; DB 4; Length 415;  
Best Local Similarity 26.2%; Pred. No. 6.7e-23;  
Matches 113; Conservative 70; Mismatches 176; Indels 72; Gaps 18;

QY 255 VGHALHEERV-----BRPPMPPTF---LYTQSWEDPEPDMVMEINPKDVTLTSG 304  
DB 16 VGVKAVYONRALS KAGISER--LFAFLPSGLVYPOIWEDEPDMVMEINPKDVTLTSG 73  
QY 305 GCNALMLLVQAGQVVSVCNPAQSALLLEKVAIQOLEFE--DVWQLFGEGVHPRIEELY 363  
DB 74 GCNILAYLTRSPARIDAVDLNAHIALNMKLEAVRPLSQGDLPRFFGAADTSHNSQAY 133  
QY 364 EKKLAPFLSQTSHNFWKRLW-----YFOHGLYYQGGMGKLCWLQCLAVVLG-----411  
DB 134 DRFIAPHLDPVSRHYWERNRGRRIAVFDRNFYQTLGLG-----LFTAMGHRTAK 185  
QY 412 -LGKTVKRLANAPTMEQORLWDSNMLIHFKVNGPKPLWLFVKFVSLVLFNKAVLWFGG 470  
DB 186 FFGVNPAPHMWEARNIGEQRRFNEELAPVFDK---KLLKWATSRKASL-----FGL 233  
QY 471 GVPKQY--ALIKADGIPIENTYAR-----TMDGVAENSHVRKQNYFYNYNCLTGKFLRDN- 523  
DB 234 GIPPAQYDSLITSGDGTMSVLKARLEKACDFPLEN-----NYFAWQAFARRYNPGE 287  
QY 524 --CPTYLREAAAFATLKSQGVVDNLTVSTNFFMBEL---KARTYTKVILMDHVDWLDMPVAN 578

DB 288 AALPAYLEKQNTETIR-GNIDRVAIHANLIEFLAGKADGTVDRLFLLDAQDMTDDQLN 346  
QY 579 ELAECLAKQVAPGGIWIWSA---SLSP--PYAELIQKAGFVRCIRRTATQGYMDRVNMY 633  
DB 347 ALWSEISRSTASAGARVIFRTAAEPSLLPGRVSTSLDDQWDYQDEASREFSA--RDRSAIY 404  
QY 634 SSFYMARRKGA 644  
DB 405 GGFHLYVVKETA 415

## RESULT 10

US-10-118-495-33  
; Sequence 33, Application US/10118495  
; Publication No. US20030074688A1  
; GENERAL INFORMATION:  
; APPLICANT: Benning, Christoph  
; APPLICANT: Riekhof, Wayne  
; APPLICANT: Klug, Rouven  
; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids  
; FILE REFERENCE: MSU-06897  
; CURRENT APPLICATION NUMBER: US/10/118,495  
; CURRENT FILING DATE: 2002-04-08  
; PRIOR FILING DATE: 2001-04-13  
; NUMBER OF SEQ ID NOS: 42  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 33  
; LENGTH: 416  
; TYPE: PRT  
; ORGANISM: Sinorhizobium meliloti  
US-10-118-495-33

Query Match 8.4%; Score 292.5; DB 20; Length 416;  
Best Local Similarity 26.0%; Pred. No. 5.1e-20;  
Matches 102; Conservative 71; Mismatches 174; Indels 45; Gaps 16;

QY 275 LYTQSWEDPEPDMVMEINPKDVTLTSGGCNALMLLVQAGQVVSVCNPAQSALLLE 334  
DB 45 VYPOIWEDEPDMVMEINPKDVTLTSGGCNALMLLVQAGQVVSVCNPAQSALLLE 104  
QY 335 KKVAVIQQ--EFEDVWQLFG--EGVHPRIEELYKELAPFLSQTSHNFW--KRLWY 385  
DB 105 KLSAFRHLPSHKDVRFLAVEGTRTN--GOAYDVFLAPKLDPATRAYMNGRDLTGRRIGV 163  
QY 386 FOHGLYYQGGMGKLCWLQCLAVVLGKTVKRLANAPTMEQORLWDSNMLIHFKVNGP 445  
DB 164 FGRNVTRTGLGRFITSASHALRLHGINP--EDFVVARSKRQORQFDDKLAFLP----E 217  
QY 446 KPLV-WLFVKFVSLVLFNKAVLWFGGVPGKQVALIKADGIPIENYIARTMDGVAENSHV 504  
DB 218 RPVIRWITSKSSL-----FGLGIPPOQFDELAS--LSREKSVAAVLRNRLEKLTJC 266  
QY 505 R---KQNYFYNYNCLTGKFLRDN---CPTYLREAAAFATLKSQGVVDNLTVSTNFFMBEL--- 555  
DB 267 HFPLRDNYPAWQAFARRYPRPDEGELPPYLOASRYEARDN--AERYEVVHASFTLLAGK 325  
QY 556 KARTYTKVILMDHVDWLDMPVANELAECLAKQVAPGGIWIWS--ASLSPP--YAEIQL 610  
DB 326 PAASVDVRYVLLDAQDMTDDQLNMLTETIRTAGAVVIFRTAAEASILPGLRSLTLLD 385  
QY 611 KAGFVRCIRRTATQGYMDRVNMYSSFYMARRK 642  
DB 386 QWYDAETSMRL--GAEDRSAYGGPHIYRKK 415

## RESULT 11

US-10-620-914-33  
; Sequence 33, Application US/10620914  
; Publication No. US20040093639A1  
; GENERAL INFORMATION:  
; APPLICANT: Benning, Christoph

APPLICANT: Riekhof, Wayne  
APPLICANT: Klug, Rouven  
TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids  
FILE REFERENCE: MSU-07769  
CURRENT APPLICATION NUMBER: US/10/620,914  
CURRENT FILING DATE: 2003-07-16  
PRIOR APPLICATION NUMBER: 10/118,495  
PRIOR FILING DATE: 2002-04-08  
NUMBER OF SEQ ID NOS: 52  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 33  
LENGTH: 416  
TYPE: PRT  
ORGANISM: Sinorhizobium meliloti  
US-10-620-914-33

Query Match 8.4%; Score 292.5; DB 4; Length 416;  
Best Local Similarity 26.0%; Pred. No. 5.1e-20;  
Matches 102; Conservative 71; Mismatches 174; Indels 45; Gaps 16;  
QY 275 LYTSQWEDPEPDMVEINPKDVTLTSGGNAALLVQAGQVSVDCNPAQSALLEL 334  
DB 45 VYQIWEDEPDMVEINPKDVTLTSGGNAALLVQAGQVSVDCNPAQSALLEL 104  
QY 335 KKVAIQOL-EFEDVWOLFQ-EGVHPRIEELYKKLAPFLSQTSHNFWSS-----KRLWY 385  
DB 105 KLSAFLPSHKDQVDFLAVEGTRTN-GQAYDFVFLAPKLPDPAATRAYNGRDLTGRRIGV 163  
QY 386 FQHGLYYQGGMGKLCWVLOCLAVVLGLGKTVKRLANAPTMBEQRRLWDSNMLHFKVNGP 445  
DB 164 FGRNVYRTGLGRFISASHALARLHGPN-EDFVKARSMEQRQPFDDKLAPLF-----E 217  
QY 446 KPLV-WLFVKFVSLVLFNKAFLWFGGVGPKQYALIKADGIPENTYIARTMDGVAENSHV 504  
DB 218 RPVIRMITRKSSL-----FGLGIPQPFDELAS--LSREKSVAAVLRNLEKJTC 266  
QY 505 R---KQNYFYNCLTGKFLRDN---CPTYLREAAFAATLKGVDNLTSTNPFMBEL--- 555  
DB 267 HFLRDNFYAQAQFARYPRPDGELPPYLOASRYEAIRDN-AERVEVHHASFTELLAGK 325  
QY 556 KARTYTKVILMDHVDMPVANELACLAKQVAPGIVWRS---ASLSPP--YAEILQ 610  
DB 326 PAASVDYVLLDAQDQWNTDQQLNDLWTEITRTADAGAVVIFRTAAEASILPGLRLSTLLD 385  
QY 611 KAGFDVRCIRTRATQGYMDRVNMYSSFYMARCK 642  
DB 386 QWYDAETSMLR--GAEDRSAYGGFHYRKK 415

RESULT 12  
US-10-684-141-58  
Sequence 58, Application US/10694141  
Publication No. US20050003536A1  
GENERAL INFORMATION:  
APPLICANT: Furusawa, Mitsuru  
TITLE OF INVENTION: METHOD AND SYSTEM FOR RAPIDLY CONFERRING  
TITLE OF INVENTION: A DESIRED TRAIT TO AN ORGANISM  
FILE REFERENCE: 690116.401  
CURRENT APPLICATION NUMBER: US/10/684,141  
CURRENT FILING DATE: 2003-10-10  
NUMBER OF SEQ ID NOS: 66  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 58  
LENGTH: 2284  
TYPE: PRT  
ORGANISM: Mus musculus  
US-10-684-141-58  
Query Match 3.6%; Score 123; DB 5; Length 2284;  
Best Local Similarity 20.8%; Pred. No. 0.094;  
Matches 141; Conservative 79; Mismatches 217; Indels 240; Gaps 35;  
QY 107 HSLCEVAKKAKAK-GWKNQVQVVEADACQAPPEGTATLITFSYS----- 150

DB 422 HNL-----KAAAKAGLYDPVELDPEDMCRMA-TEQPTLATYVSDAVATYLYMKYVHP 476  
QY 151 -----LTMIPFFH-NVI-----DQACSVLSODGLVGVA 177  
DB 477 FIFALCTIIPMEPDEVLRKSGTGLCALLMWQAFAHANIIFPNKQEQEFNKLTDDGHVLD 536  
QY 178 DFVVG-----KYDLP--LRQMPWSRFRFFWRSIFDIDNIDIGPERRAYLEQ--KL 223  
DB 537 EYVVGHVHVALBSGVPRSDIPCFRFRNPAAPFELLQORVEKTMRHAETEEEEKVPEQATNF 596  
QY 224 ERVWEQ--NTQGSIPYVWLRAPYVYVIMGRPLSVGHALHEERVERP-----PMFPPT 273  
DB 597 QEVCEQIKTKLTSKQV-----NRIECPLIYHLDVGAMYVNI 634  
QY 274 FLYTSQWEDPEPDMVEINPKDVTLTSGGNA-----ALNLLVQAGQVSVDCN 325  
DB 635 ILTNR-----LQSAIVDEATCAACDFNKPAGSCQKQKMAWQWGEFM----- 676  
QY 326 PAQSALLELKVAIQOLEFEDVWOLFEGVHPRIEL-----YEK-KLAPF----- 370  
DB 677 PASRS--EYHRIQ-HOLESEKPPPLFPEGPARAFHLSREEQAKYKRLADYCRKAYKK 733  
QY 371 -----LSQTSNFWKRLWYFQHGLYYQGGMGKLCWVLOCLAVVLGLGKTVKR 418  
DB 734 IHVTKVEERLTTICQRENSFYVDTVRAFRDRRYEPKGLHKWKKLSAAVEVDGASEVKR 793  
QY 419 LANAPTEBQRRLWDSNMLH-----FVKGPKPLVWLFVRFVSLVLFNKAVLWPF 468  
DB 794 CKN-----MEILYDLSQLAHKILNSFYGVNMRKAR--WYSMEMAGIVCPT----- 838  
QY 469 GGGVPGKQYALIKADGIP-----ENYIARTMDGVAENSHVKNQYFY 512  
DB 839 GANITQARELEIQGRPLELTDGICWVLPNSFPENFIKT-----TNAKKPKLITISYP 893  
QY 513 NCLTGKFLRDNCTY-----LREAAFAATLKGVDNLTSTNPFMBELKARTYTKVILMDH 568  
DB 894 GAMLINWKEGFTNHQYQELTEPSLT-----VTHSENSIFFEVDG-PYLAMIL--- 942  
QY 569 VDMLDMPVANELACLAKQVAPGIVWRSASLSPYAEILQKAGFDVRCIRATQGYMD 628  
DB 943 -----PASKEGKKLKRYA-----VFNEGDSL-----KGFEVK-----RRGELQ 980  
QY 629 RVNMY--SSFYMARCKGA 644  
DB 981 LKIFQSSVFEAFKGS 997

RESULT 13  
US-10-810-486-58  
Sequence 58, Application US/10810486  
Publication No. US20050054597A1  
GENERAL INFORMATION:  
APPLICANT: Furusawa, Mitsuru  
TITLE OF INVENTION: METHOD AND SYSTEM FOR RAPIDLY CONFERRING  
TITLE OF INVENTION: A DESIRED TRAIT TO AN ORGANISM  
FILE REFERENCE: 690116.401C1  
CURRENT APPLICATION NUMBER: US/10/810,486  
CURRENT FILING DATE: 2004-03-26  
PRIOR APPLICATION NUMBER: US 10/684,141  
PRIOR FILING DATE: 2003-10-10  
PRIOR APPLICATION NUMBER: JP 2003-092898  
PRIOR FILING DATE: 2003-03-28  
NUMBER OF SEQ ID NOS: 95  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 58  
LENGTH: 2284  
TYPE: PRT  
ORGANISM: Mus musculus  
US-10-810-486-58  
Query Match 3.6%; Score 123; DB 5; Length 2284;  
Best Local Similarity 20.8%; Pred. No. 0.094;



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: March 14, 2006, 02:01:27 ; Search time 9057 Seconds  
(without alignments)  
4066.971 Million cell updates/sec

Title: US-10-620-914-45  
Perfect score: 3463  
Sequence: 1 MSGRDCRPASYTKNFKNSLE.....RVNMYSSFYMARKGAKGN 648

Scoring table:

BLOSUM62	Xgapop 10.0 , Xgapext 0.5
-Q-/abes/ABSSWEB	Ygapop 10.0 , Ygapext 0.5
-Q-/abes/ABSSWEB	Ygapop 6.0 , Ygapext 7.0
Delop 6.0 , Delext 7.0	

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

```
-MODEL=frame+ p2n.model -DEV=xlh
-Q-/abes/ABSSWEB spool/US10620914/runat_13032006_102040_6725/app_query.fasta_1
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biom62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abes05h
-USER=US10620914 @CGN 1.1 4939 @runat_13032006_102040_6725 -NCPU=6 -ICPU=3
-NO MWAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7
```

## Database :

```
GenEmbl:*
1: gb.ba:*
2: gb.in:*
3: gb.env:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pr:*
9: gb.ro:*
10: gb.sts:*
11: gb.sy:*
12: gb.un:*
13: gb.vi:*
14: gb.htg:*
15: gb.pl:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	3441	99.4	2001	15	AY656806	Chlamydom
2	913.5	26.4	31065	15	AC152119	Pichia st
3	910	26.3	110000	15	CR382131_12	Continuation (13 o

4	890.5	25.7	110000	15	CR382126_20	Continuation (21 o
c 5	864.5	25.0	110000	15	CR382138_20	Continuation (21 o
6	804	23.2	75337	15	BX842626	Neurospor
7	724	20.9	6962	15	AJ878596	Rhizopus
8	633	18.3	110000	15	AE017345_05	Continuation (6 of
c 9	460	13.3	311050	1	BX294133	Pirellula
10	395	11.4	349260	1	BX527595	Rhodosphe
11	379.5	11.0	3045	1	AF329857	Rhodosphe
12	352	10.2	933	6	AE551999	Sequence
13	336.5	9.7	10129	1	AE008128	Agrobacte
14	336.5	9.7	10134	1	AE009162	Agrobacte
15	329	9.5	110000	1	BA000012_13	Continuation (14 o
16	292.5	8.4	294800	1	SME591789	Continuation (14 o
c 17	277	8.0	86289	14	CR954215_4	Continuation (5 of
c 18	243.5	7.0	756	10	PM8H7B	AL684262 Penicilli
c 19	232.5	6.7	661	10	PM11C4B	AL684149 Penicilli
c 20	178	5.1	344249	1	BX842654	Bdellovib
21	159	4.6	110000	1	AP006618_16	Continuation (17 o
22	158	4.6	866	6	AR557101	Continuation (17 o
c 23	138.5	4.0	8832	6	AR619589	AR57101 Sequence
c 24	137	4.0	725	10	PM3B11G	AL684886 Penicilli
25	132.5	3.8	306358	1	AE016759	AL684755 Escherich
26	131.5	3.8	300409	1	AE016755	AL684755 Escherich
27	128	3.7	4246	6	AX702426	Sequence
28	126.5	3.7	590	6	AR627802	Sequence
29	123.5	3.6	203961	8	AC011374	AC011374 Homo sapi
30	123	3.6	7119	6	CQ893379	CQ893379 Sequence
31	123	3.6	7119	9	AF123502	AF123502 Mus muscu
32	123	3.6	340900	1	SME591791	AL591791 Sinorhizo
33	122	3.5	1604	5	CC515KDP	X15825 Quail mRNA
34	122	3.5	7102	9	BC048166	BC048166 Mus muscu
35	122	3.5	7102	9	BC063246	BC063246 Mus muscu
c 36	122	3.5	7954	1	AE010932	AE010932 Methanosa
37	121	3.5	639	6	AR348989	AR348989 Sequence
38	121	3.5	10029	1	AE013909	AE013909 Yersinia
c 39	121	3.5	210050	1	AJ414146	AJ414146 Yersinia
40	121	3.5	290510	1	AE017130	AE017130 Yersinia
41	119	3.4	5067	8	AB209902	AB209902 Homo sapi
42	119	3.4	6912	6	CQ893375	CQ893375 Sequence
43	119	3.4	6912	8	HUMDNPOLCS	L09561 Homo sapien
44	118.5	3.4	1155	6	BD242271	BD242271 Compounds
45	118.5	3.4	1155	6	AR278478	AR278478 Sequence

## ALIGNMENTS

RESULT 1	AY656806	2001 bp	mRNA	linear	PLN 11-FEB-2005
AY656806	Chlamydomonas reinhardtii betaine lipid synthase (BTAL) mRNA, complete cds.				
ACCESSION	AY656806				
VERSION	AY656806.1				
KEYWORDS	GI:50261571				
SOURCE	Chlamydomonas reinhardtii				
ORGANISM	Chlamydomonas reinhardtii				
REFERENCE	1 (bases 1 to 2001)				
AUTHORS	Riekhof, W.R., Sears, B.B. and Benning, C.				
TITLE	Annotation of Genes Involved in Glycerolipid Biosynthesis in Chlamydomonas reinhardtii: Discovery of the Betaine Lipid Synthase BTALC				
JOURNAL	Eukaryotic Cell 4 (2), 242-252 (2005)				
PUBLISHED	15701786				
REFERENCE	2 (bases 1 to 2001)				
AUTHORS	Riekhof, W.R. and Benning, C.				
TITLE	Direct Submission				
JOURNAL	Submitted (17-JUN-2004) BMB, MSU, Biochemistry 215, East Lansing, MI 48824-1319, USA				
FEATURES	Location/Qualifiers				
source	1..2001				
	/organism="Chlamydomonas reinhardtii"				





```

Db 1561 CACGTGGCAGCAGCACTACTTCTACTACAACTGCTCACCAGGCAAGTTCTCGCGGAC 1620
Qy 523 AsnCysProThrTyrLeuArgGluAlaAAlaPheAlaThrLeuLysSerGlyValValAsp 542
Db 1621 AACTGCCCCACCTACCTGCGGAGGCGGCTTCGCCACCTCAAGAGTGGCGTGGAC 1680
Qy 543 AsnLeuThrValSerThrAsnPhePheMetGluGluLeuLysAlaArgThrTyrThrLys 562
Db 1681 AACCTGACCGCTCTCCACCACTTCTTCATGGAGAGCTCAAGCGCGCACTTACACCAAG 1740
Qy 563 ValLeuLeuMetAspHisValAspTyrLeuAspMetProValAlaAsnGluLeuAlaGlu 582
Db 1741 GTGATCTGTATGACCAACGCTGACTGGCTGATATGCCCGTGGCCACGAGCTGGCGGAG 1800
Qy 583 CysLeuAlaLysGlnValAlaProGlyGlyLeuValLeuTyrArgSerAlaSerLeuSer 602
Db 1801 TGCCTGGCCAAAGCAGGTGGCGCGCGGCATCGTATCTGGCGCTCCGCTCCCTCAGC 1860
Qy 603 ProProTyrAlaGluLeuLeuGlnLysAlaGlyPheAspValArgCysIleArgArgAla 622
Db 1861 CCGCCCTACGCCGAGCTGATCCAGAGAGCGGGCTTCGACGTGGCTGCATCCCGCGGCC 1920
Qy 623 ThrGlnGlyTyrMetAspArgValAsnMetTyrSerSerPheTyrMetAlaArgArgLys 642
Db 1921 ACTCAGGCTACATGATGACCGCTCAACATGTACAGTCTCTTCTACATGGCCCCCGGAG 1980
Qy 643 GlyAlaLysLysAspAsn 648
Db 1981 GGGCCCAAGAAGGACCAAC 1998

RESULT 2
AC152119 31065 bp DNA linear PLN 30-NOV-2004
LOCUS Pichia stipitis clone JGIAHYZ-5A24, complete sequence.
DEFINITION AC152119
ACCESSION AC152119.2 GI:56121961
VERSION HTG.
KEYWORDS Pichia stipitis
SOURCE Pichia stipitis
ORGANISM Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Pichia.
REFERENCE 1 (bases 1 to 31065)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 31065)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (27-Oct-2004) Production Genomics Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA
94598-1698, USA
REFERENCE 3 (bases 1 to 31065)
AUTHORS Stanford Human Genome Center.
CONSTRM DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (30-NOV-2004) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Nov 30, 2004 this sequence version replaced gi:54654153.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated Total Number of Errors is 0.
FEATURES
source Location/Qualifiers
1..31065
/organism="Pichia stipitis"
/mol_type="genomic DNA"
/db_xref="taxon:4924"
/clone="JGIAHYZ-5A24"

ORIGIN
Alignment Scores:

```

```

Pred. No.: 1,25e-69 Length: 31065
Score: 913.50 Matches: 223
Percent Similarity: 46.5% Conservative: 113
Best Local Similarity: 30.8% Mismatches: 244
Query Match: 26.4% Indels: 143
DB: 15 Gaps: 21

US-10-620-914-45 (1-648) x AC152119 (1-31065)
Qy 40 GlySerLysGlyAspAspHisAlaAlaArgLeuGluSerPheTyrGlyProGlnAla 59
Db 20013 GCAAGAAGAGGATCAGCAACATTGG-----CTTGAGCAATTTCTACAAGAACAGGCT 20066
Qy 60 AlaAlaPhe----- 62
Db 20067 CACATCTACGACACCAACCCGCTGAGTTCTTGTGAAGGGAAGACAGGAATGTCTTCGCTT 20126
Qy 63 ---AlaAlaArgLeuAlaGluArgSerAsnLeuLeuTyrValAspLeuGlyGlyThr 81
Db 20127 GCTGCTCCATTTGCGAAGAAAGAAAGAGTGGTCTGGATCGATTTGGTGGAGGTACT 20186
Qy 82 GlyGluAsnValAspMetMetAlaAspTyrIleAspLeuAlaLys---PheLysSerIle 100
Db 20187 GGTTCCAACATCGAGTACATGGACCACTGCTCTCCATTTCGAAAACTTTAAGGCGCTC 20246
Qy 101 TyrValValAspLeuCysHisSerLeuGluValAlaLysLysLysLysAlaLys 120
Db 20247 TATTTGGTTGATTTGCTCTCTCTCTGAGTGGCTCGTGCCAGATTTCGACGCTCGT 20306
Qy 121 GlyTyrLysAsnValGlnValValGluAlaAspAlaCysGlnPheAlaProGluGly 140
Db 20307 AACTGGAAGAAATGTCAAGGTCAATTTGCTGATGCTGTGACTTTGAAATCGACTACAAG 20366
Qy 141 ThrAlaThrLeuIleThrPheSerTyrSerLeuThrMetIleProPheHisAsnVal 160
Db 20367 ACTGCTGACTTGATCAGTTTTCCTACTCTCTTCTATGATTCCACTTCAATGCTGCT 20426
Qy 161 IleAspGlnAlaCysSerTyrLeuSerGlnAspGlyLeuValGlyValAlaAspPheTyr 180
Db 20427 ATCGACATGCACTCACCAGTTGGATAAGCAGGGTATTATTGCTGCTGTAGACTTTGGT 20486
Qy 181 ValSerGlyLysTyrAspLeuProLeu----- 189
Db 20487 ATCCAG---AGTTTGGACACCTCTATGGGCGAGAAATCAACACTGTAGGAGGTTGGTCAAC 20543
Qy 190 ArgGlnMetProTyrSerArgArgPhePheTyrArgSerIlePheAspIleAspAsnIle 209
Db 20544 AGAATCATCCCTGGGTGCTCAGAACTTCTGGAGAAATTTGGTTGAAGCCGCAAGTC 20603
Qy 210 AspIleGlyProGluArgAlaTyrLeuGluGlnLysLeuGluArgValTrpGluGln 229
Db 20604 TTTTGGACTCATCGAAGAAATTTACTTGAATACAAAGTTTGAACCACTCAAGTCTCTC 20663
Qy 230 AsnThrGln-----GlySerIleProTyrValProTrpLeuArgAlaProTyr 245
Db 20664 AACACCTATAACAAGAAAGTTGGGTAAATTC-----CCTTAC 20699
Qy 246 TyrValTrpIleGly-----ArgLeuProSerValGlyHisAlaLeuHisGlu 261
Db 20700 TACATCTGGATTTGGCTGTGACCAAGTCCAAAGTCCGAAATATCTTACCGCTCTCACTGC 20759
Qy 262 GluArgValGluArgProMetPheProThr----- 273
Db 20760 TTGGCTACTGAATCTCTTACTTGGCACCTTCTACCACCTCCCATTTACCACCTGCTCTCAA 20819
Qy 273 ----- 273
Db 20820 GACGTTCTTATCAGTAAGGGCCCATGAAGCTGCGCTTGACGAATTGCAAAAGAACTTACCT 20879
Qy 273 ----- 273
Db 20880 TTTCTCTTTCTACTACTACCGCGGCAATCTGGAGAGTGTACTATGAAGAGTTGAACCAG 20939

```

```
QY 274 -----PheLeuTyrThrGlnSerTrpGluAspProGluPro 285
Db 20940 CATTACGACAGTTCAAGAACCAATACATCTACGCCTTCACATGGGAGAGATCCACGTGAA 20999
QY 286 AspMetGluValMetGluIleAsnProLysAspThrValLeuThrLeuThrSerGlyGly 305
Db 21000 GATCACAATATCTCAAGTTCAACAGCAATGACACAGTTTTAGCCATTACTTCTGCGGGT 21059
QY 306 CysAsnAlaLeuAsn-----LeuLeuValGlnGlyAlaGlyGlnValValSerValAsp 323
Db 21060 GACAACATCTTATCTTATGCTACCTTACCGGACCCCAAGAGAAATCCATGCTGTAGAC 21119
QY 324 CysAsnProAlaGlnSerAlaLeuLeuGluLeuLysLysValAlaIleGlnGlnLeuGlu 343
Db 21120 TTAACCCATCTCAAAACCACTGTTGGAGTTGAAGTTGCAGCTTTCAGATCTCTTGACA 21179
QY 344 PheGluAspValTrpGlnLeuPheGlyGlyValHisProArgIleGluGluLeuTyr 363
Db 21180 CACCAACAGATCTGGCAGATGTTGGTGAAGGTAAGATCGAAACTTCACCGACTTGTG 21239
QY 364 GluLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrpSerLysArg--- 382
Db 21240 ATCGACAAGTTGTCACACATCTCTTCCAAATGCCCTTCAGTACTGGATGGACAGAGGA 21299
QY 383 ---LeuTrpTyrPheGlnHisGlyLeuTyrTyrGlnGlyGlyMetGlyLysLeuCysTrp 401
Db 21300 CACAAGTGTCTCCAGAACGACACTTAC-----CACACTGGTCTCTAGTAGTGG 21350
QY 402 ValLeuGlnCysLeuAlaVal-----ValLeuGlyLeuGlyLysThrValLysArg 418
Db 21351 GCTCTCAGATCGCCAGTAGTACACTTTCAGGCGCTTGTGTCTCAATAAGATATCAAGCAG 21410
QY 419 LeuAlaAsnAlaProThrMetGluGluGlnArgLeuTrpAspSerAsnMetLeuIle 438
Db 21411 TTGTGCGAAGCAGCTACTATGGACGACAGATGAAGATCTGGGAAGAGAGA----- 21461
QY 439 HisPheValLysAsnGlyProLysProLeuValTrpLeuPheValLysPheVal---Ser 457
Db 21462 -----GTGAAGAAGTCT-----TTGTCAACTCAGTAGTGGCTACT 21497
QY 458 LeuValLeuPheAsnLysAlaValLeuTrpPheGlyGlyValProGlyLysGlnTyr 477
Db 21498 TTATTGTTGGAACCCACTCTTTTGTGGAAGCTCTTGAGAGTCTCCAGCTAATCAGGCC 21557
QY 478 AlaLeuIleLysAlaAspGlyIleProIleGluAsnTyrIleAlaArgThrMetAspGly 497
Db 21558 TCCATGATG-----GGAGAATCGCTCTTCAAGTATGTGTGCGACAAATAGACCT 21608
QY 498 ValAlaGluAsnSerHisValArgLysGlnAsnTyrPheTyrTyrAsnCysLeuThrGly 517
Db 21609 CTCTTGAAGAGATCTTGATTTCAAGAGACANATTTTCTACTACTTGTGCTTGATGAAC 21668
QY 518 LysPheLeuArgAspAsnCysProThrTyrLeuArgGluAlaAlaPheAlaThrLeu--- 536
Db 21669 AAGTACGCTACAACTGTCCGACTCTTCAAGGCTCAAGGCTCAAGAGATTGACG 21728
QY 537 -----LysSerGlyValValAspAsnLeuThrValSerThrAsnPhe 550
Db 21729 GCCCAGACTACTCCCAAGAGAAACCCATTGGACACATCAGATTCATCTACTGACTTT 21788
QY 551 PheMetGlu-----GluLeuLysAlaArgThrTyrThrLysValIleLeuMetAsp 567
Db 21789 TTGAACGAGCTGTTGCCAGATTGAGCAGAGAGTCTGTGACTATCGCAATTTATGAGAC 21848
QY 568 HisValAspTrpLeu-----AspMetProValAlaAsnGluAlaGluCysLeuAla 585
Db 21849 CACATGATGTTGTTGATCATGACTCCCAAGGATGTCAACGATGATGATTCGTTCTTAAG 21908
QY 586 LysGlnValAlaProGlyGlyIleValIleTrpArgSerAlaSerLeuSerProProTyr 605
Db 21909 GAAAGTCTTTCGACGGTGGTAGATCTTGTGAGATCTGCTCTCAACCCCATGTATC 21968
QY 606 AlaGluLeuIleGlnLysAlaGlyPheAspValArgCysIleArg---ArgAlaThrGln 624
```

```
Db 21969 ATCAGCACCTTTGAGTCTATGGCTTTAGTTGTAAGCTGTCTGTCAAAATCTCAGGT 22028
QY 625 GlyTyrMetAspArgValAsnMetTyrSerSerPheTyrMetAlaArgArgLysGlyAla 644
Db 22029 GAATCTATTGACAGGGTCAACATGTATGCTTCTACTTGGGTATGCACCAAGCTTCCAACC 22088
QY 645 LysLysAsp 647
Db 22089 AAGGAAGAC 22097
```

## RESULT 3

CR382131\_12

## WPCOMMENT

Sequence split into 43 fragments LOCUS CR382131 Accession CR382131.

Fragment Name Begin End

```
CR382131_00 1 110000
CR382131_01 100001 210000
CR382131_02 200001 310000
CR382131_03 300001 410000
CR382131_04 400001 510000
CR382131_05 500001 610000
CR382131_06 600001 710000
CR382131_07 700001 810000
CR382131_08 800001 910000
CR382131_09 900001 1010000
CR382131_10 1000001 1110000
CR382131_11 1100001 1210000
CR382131_12 1200001 1310000
CR382131_13 1300001 1410000
CR382131_14 1400001 1510000
CR382131_15 1500001 1610000
CR382131_16 1600001 1710000
CR382131_17 1700001 1810000
CR382131_18 1800001 1910000
CR382131_19 1900001 2010000
CR382131_20 2000001 2110000
CR382131_21 2100001 2210000
CR382131_22 2200001 2310000
CR382131_23 2300001 2410000
CR382131_24 2400001 2510000
CR382131_25 2500001 2610000
CR382131_26 2600001 2710000
CR382131_27 2700001 2810000
CR382131_28 2800001 2910000
CR382131_29 2900001 3010000
CR382131_30 3000001 3110000
CR382131_31 3100001 3210000
CR382131_32 3200001 3310000
CR382131_33 3300001 3410000
CR382131_34 3400001 3510000
CR382131_35 3500001 3610000
CR382131_36 3600001 3710000
CR382131_37 3700001 3810000
CR382131_38 3800001 3910000
CR382131_39 3900001 4010000
CR382131_40 4000001 4110000
CR382131_41 4100001 4210000
CR382131_42 4200001 4224103
```

Continuation (13 of 43) of CR382131 from base 1200001 (CR382131 Yarrowia lipolytica chrom

## Alignment Scores:

```
Pred. No.: 1-29e-68 Length: 110000
Score: 910.00 Matches: 226
Percent Similarity: 48.5% Conservatives: 116
Best Local Similarity: 32.1% Mismatches: 231
Query Match: 26.3% Indels: 132
DB: 15 Gaps: 22
```

US-10-620-914-45 (1-648) x CR382131\_12 (1-110000)

QY 51 LeuGluSerPheTyrGlyProGlnAla-----

|||||

Db 59694 CTGGAGTCTTCTTACAAAGGCCGAGCCCAACATTTACGACCGAACCAGGACCATCTGTGCTC 59753  
Qy 60 -----AlaalaPheAlaalaArgLeuAlaGluArgSerAsnLeu 72  
Db 59754 AAGGCCGAGAGCAGGCTCTGGCCACCACCGCCGCGAGCTTCGAAGAAGCATGATCTC 59813  
Qy 73 IleTrpValAspLeuGlyGlyThrGlyGluAsnValAsp---MetMetAlaAspTyr 91  
Db 59814 GTGTGGTTCGACATTTGGAGGAGAACCGCGTGGAAACATTGAGCAGCATGACCAACTAC 59873  
Qy 92 IleAspLeuAlaLysPheLysSerIleTyrValValAspLeuCysHisSerLeuCysGlu 111  
Db 59874 CTTCCCATTTCTGCTTCAAGGCATCTACTGTTGATCTGAGTCTCTCTGTGCGAG 59933  
Qy 112 ValAlaLysLysLysAlaLysAlaLysGlyTrpLysAsnValGlnValValGluAlaAsp 131  
Db 59934 GTGGCCGAAGCGATTTCGCCGAGAGAACTGGAGAACTGCAAGTCTGTGTCATCGAC 59993  
Qy 132 AlaCysGlnPheAlaProProGluGlyThrAlaThr---LeuIleThrPheSerTyrSer 150  
Db 59994 GCTGCCGATTTCATCTCCCGCGTGGATCTCTCGTTGATCTCTTCCACCATGCTACTCT 60053  
Qy 151 LeuThrMetIleProProPheHisAsnValIleAspGlnAlaCysSerTyrLeuSerGln 170  
Db 60054 CTGTATGATCCCCACCTACTATGCGGTATGATCGACTCTCCACCTGCTTGGCAAG 60113  
Qy 171 AspGlyLeuValGlyValAlaAspPheTyrValSerGlyLysTyrAspLeuPro----- 188  
Db 60114 GACGGTCTGTCTACTGTCATGCTGCTTACGTCCTCAATCCATCCACCTTGACCGCCAG 60173  
Qy 189 -----LeuArgGlnMetProTrpSerArgArgPhePheTrp 200  
Db 60174 TCTACTACATGGCGCGAGCTGCTCCGCGATGTCACTGGTCTTCCCGAACTTTCTGG 60233  
Qy 201 ArgSerIlePheAspIleAspAsnIleAspIleGlyProGluArgArgAlaTyrLeuGlu 220  
Db 60234 CGACTGTGGTTCGAGTTGACCGGGTCTACTCGACTCTGCCGAGAGACTATCTCGAG 60293  
Qy 221 GlnLys-----LeuGluArgValTrpGluGlnAsnThrGln---GlySerIlePro 236  
Db 60294 TACCGATTCCGAACCATCAGTCACTGCTGCGGAACACCGACTCGGAGGAACTCCCC 60353  
Qy 237 TyrValProTrpLeu----- 241  
Db 60354 TACTACTACTGTGATTTGGATGTGCAAGGACCGATCCACCAACATTTCTCCAGCGCCAAC 60413  
Qy 242 -----ArgAlaProTyrTyr-----ValTrpIleGlyArgLeu 252  
Db 60414 GCTCTGTGTACCGAGTCTCCCTACCTGTCCCCCAGAACAGAGCTTGTGTGGAGACGAG 60473  
Qy 253 ProSerValGlyHisAlaLeu----- 259  
Db 60474 GACGCGCTAACACACTCTGGGTATCCGATCCAGGGCTACGATGCCGCTCTCGTCNAC 60533  
Qy 259 ----- 259  
Db 60534 ATGCAGCGAACTTCCCGCTCCCTCATTTCTTACCAGACCGAAATCTGGCGAATCTTC 60593  
Qy 260 HisGluGluArgValGluArgProProMetPheProThrPheLeuTyrThrGlnSer 279  
Db 60594 TACAACGAGGAGTGTCCCAAGTACCACAGTTTGGCCAACAGTACATACGCTTCCAC 60653  
Qy 280 TrpGluAspProGluProAspMetGluValMetGluIleAsnProLysAspThrValLeu 299  
Db 60654 TGGGAGGACCCCGAGGAGACAGAACATTTCTACAGTTTAAGCCCGAGCATACCGTCTT 60713  
Qy 300 ThrLeuThrSerGlyCysAsnAlaLeuAsnLeuValGlnGlyAla-----Gly 317  
Db 60714 GCCATCACTCTGCGGTGACAACTTCTGTCTTACGCTCCCATGAGCGCTCCCGCCNAG 60773  
Qy 318 GlnValValSerValAspCysAsnProAlaGlnSerAlaLeuLeuLeuLysVal 337  
Db 60774 CGAATCAATTGTCTGACCTCAACCTTGCAGAACCATCTGCTGAGCTCAAGCTGGCC 60833

## RESULT 4

CR382126\_20

## WPCOMMENT

Sequence split into 26 fragments LOCUS CR382126 Accession CR382126

Qy 338 AlaIleGlnGlnLeuGluPheGluAspValTrpGlnLeuPheGlyGluGlyValHisPro 357  
Db 60834 TGTCTCGAGTGTCTTCCCTTTGAGGACATGTGGAAGCTGTTTGGTGAAGGCAAGCACCCC 60893  
Qy 358 ArgIleGluGlnLeuTyrGluLysLysLeuAlaProPheLeuSerGlnThrSerHisAsn 377  
Db 60894 AAGTTCGAAGAGCTGTGACCAAGCTGGCCCCCACCCTGTCTTCCCAAGCTTCCAG 60953  
Qy 378 PheTrpSerLysArgLeuTrpPheGlnHisGlyLeuTyrTyrGlnGlyGlyMet--- 396  
Db 60954 TACTGGCAGACGCA-----GGCCAGCCTCTTTACCGGAAGGCTGTGTTT 61001  
Qy 397 -----GlyLysLeuCysTrpValLeuGlnCys-----LeuAlaValValLeuGly 411  
Db 61002 GATACCGGCTTCTCCGATGGGCCATTTCGACTTGGCATTTGGTCTTTCGCAATTTCTGTT 61061  
Qy 412 LeuGlyLysThrValLysArgLeuAlaAsnAlaProThrMetGluGlnGlnArgArgLeu 431  
Db 61062 GTCTCCCCCACATTAGCGCTCTGTGCGAGGCCAAGACTCTCGACGAACAGTGGAGCGTC 61121  
Qy 432 TrpAspSerAsnMetLeuIleHisPheValLysAsnGlyPro-----LysProLeu 448  
Db 61122 TGGAGAAGTCCCTT-----CGACCTGTCTGTCCAAACCCCTT 61160  
Qy 449 ValTrpLeuPheValLysPheValSerLeuValPheAsnLysAlaValLeuTrpPhe 468  
Db 61161 GTC-----GCCAAGTGTCTTGGTGGCAACCCCATTTTCTGTGGAAG 61202  
Qy 469 GlyGlyValProGlyLysGlnTyrAlaLeuIleLysAlaAspGlyIleProIleGlu 488  
Db 61203 GCTCTCGGTGTCTTCCCTGAGCAGCTCTATGATCGAGGCGGTATGCTC----- 61253  
Qy 489 AsnTyrIleAlaArgThrMetAspGlyValAlaGluAsnSerHisValArgLysGlnAsn 508  
Db 61254 AGTTTGTCTATCGACACTTTCGAGCCCATCATCAGCGATCGCTCATTTCCGAGCAGCAC 61313  
Qy 509 TyrPheTyrTyrAsnCysLeuThrGlyLysPheLeuArgAspAsnCysProThrTyrLeu 528  
Db 61314 TACTTCTACTCTGTCTCAAGGCTGTCTAGCGTCCCAACAACTGTCCCGACTTCTCTG 61373  
Qy 529 ArgGluAlaAlaPheAlaThrLeu-----LysSerGlyValValAspAsnLeuThrVal 546  
Db 61374 ACCAAGCAGCGCCACCCCACTTGGCCAAAGCGAAGGAGCTTTGACGGAATCCGAATC 61433  
Qy 547 SerThr-----AsnPhePheMetGluLeuLysAlaArgThrTyrThrLysVal 563  
Db 61434 CACACCGAGATCAACGAGGTGGTCAAGCGACTCAACCCGATCAGTCAACACGCCC 61493  
Qy 564 IleLeuMetAspHisValAspTrpLeu-----AspMetProValAlaAsnGluLeuAla 581  
Db 61494 ATCGTTATGGACCATGAGTGTTCCTCCCAAGGACGGTAACGCCCTCGAGAGGAGATC 61553  
Qy 582 GluCysLeuAlaLysGlnValAlaProGlyGlyIleValIleTrpArgSerAlaSerLeu 601  
Db 61554 AAGTCTCTCCACCATCTCTTACCGAAGTGGAAACGTTATGTTCGATCAGCTCCAG 61613  
Qy 602 SerProTyrAlaGluLeuIleGlnLysAlaGlyPheAspValArg-----CysIle 619  
Db 61614 AAGCCTGTGTACTCTAGCTCTATGAGGAGGAGGCTTACCACTAGACCCGCTGCCATC 61673  
Qy 620 ArgArgAlaThrGlnGlyTyrMetAspArgValAsnMetTyrSerSerPheTyrMetAla 639  
Db 61674 CGAGAATCTGGCACTTCC---ATTGACCGAGTCAACATGTACGCTTCCACTTCCGCTGTG 61730  
Qy 640 ArgArgLysGlyAla 644  
Db 61731 ACCAAGAGGAGGCC 61745

Fragment Name	Begin	End	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
CR382126_00	1	110000	7,01e-67	890,50	46.6%	33.2%	25.7%	15
CR382126_01	100001	210000						
CR382126_02	200001	310000						
CR382126_03	300001	410000						
CR382126_04	400001	510000						
CR382126_05	500001	610000						
CR382126_06	600001	710000						
CR382126_07	700001	810000						
CR382126_08	800001	910000						
CR382126_09	900001	1010000						
CR382126_10	1000001	1110000						
CR382126_11	1100001	1210000						
CR382126_12	1200001	1310000						
CR382126_13	1300001	1410000						
CR382126_14	1400001	1510000						
CR382126_15	1500001	1610000						
CR382126_16	1600001	1710000						
CR382126_17	1700001	1810000						
CR382126_18	1800001	1910000						
CR382126_19	1900001	2010000						
CR382126_20	2000001	2110000						
CR382126_21	2100001	2210000						
CR382126_22	2200001	2310000						
CR382126_23	2300001	2410000						
CR382126_24	2400001	2510000						
CR382126_25	2500001	2602197						

Continuation (21 of 26) from base 2000001 (CR382126 from base 2000001 Kluveromyces lactis str

Alignment Scores:	Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	DB:
51	LeuGluSerPheTyrGlyProGlnAlaAlaPhe	110000	46.6%	33.2%	25.7%	15
75238	TTAGACAATTTCTATAAATCCCAAGCAAACTCTACGATCGTACAAAGGGTGTCTTACTC	75297				
63	-----AlaAlaArgLeuAlaGluArgSerAsnLeu	72				
75298	CAAGTCGTGAACAGCCTTAATATCTCTCTCACCTCTCTGAAAGAAAGGTAAT	75357				
73	IleTrpValAspLeuGlyGlyThrGlyGluAsnValAspMetMetAlaAspTyrIle	92				
75358	GTCTGGATTGATGTCGGGGTGGGACTGGTTTCAATATCTCCCAATATGGCGCTTTAAACA	75417				
93	AspLeu---AlaLysPheLysSerIleTyrValValAspLeuCysHisSerLeuCysGlu	111				
75418	AACTTGGATACACAGTTTACAAAATTTACTTGATGATCTCTCCATCGCTATGTGAG	75477				
112	ValAlaLysLysLysAlaLysAlaLysGlyTrpLysAsnValGlnValValGluAlaAsp	131				
75478	GTTCAGAAAGACGTTGTAAGACATGATGGAAGAAATGGAAGAGATCTGTGGAGAT	75537				
132	AlaCysGlnPheAlaProProGluGlyThrAlaThrLeuIleThrPheSerTyrSerLeu	151				
75538	GCTTCGATTTGAATACCAAGAGGAATCAGTCAGTTGATCACAATTTCTTATTCATTTG	75597				
152	ThrMetIleProProPheHisAsnValIleAspGlnAlaCysSerTyrLeu---SerGln	170				
75598	AGCATGATACCAAGTTTCTTCGCGCCATTTGATCAGCAGTATCTTTATTGGATGCTAAG	75657				
171	AspGlyLeuValGlyValAlaAspPheTyrValSerGlyLysTyr-----	185				
75658	AATGATATCTTCGTGTGTTGATTTTGGTGCTACTAATGAATCTATGCTAGTCGGAAGA	75717				
186	-----AspLeuProLeuArgGlnMetProTrpSerArgArgPheTyr	200				

75718	ACCAACACATTTGGTGGTCTAGTTAAACAGACATATACCTTGGCTATTCCGTACATTCGG	75777
201	ArgSerIlePheAspIleAspAsnIleAspIleGlyProGluArgArgAlaTyrLeuGlu	220
75778	AGATTGGTGTGAATTCGACAAAGTGTTCCTGATCCAGCTAGAGAGATATCTTGAA	75837
221	GlnLysLeuGluArgValTrpGluGlnAsnThrGln-----GlySerIlePro	236
75838	TATAGATTCGGGACAAATCAATCAATTTGTTATTAACATCAACAGCTTGTGTAATTTCCC	75897
237	TyrValProTrpLeu-----	241
75898	TACTACATTTGGTGGTTCGCAATAAAGACCACCAACACATCTTCAAGCTAGATTTGTT	75957
242	-----ArgAlaProTyrTyrValTyrIleGlyArg-----	251
75958	GAATTCGCACCAACATCTCCATACCTCGCACCTATAACACAGCTCCTCTCAAGCT	76017
252	-----Leu	252
76018	CAGCAATGACCAAGCTATGATTGCAGCACTAGAAAAATTTCAAAAAAGGTTTACCATA	76077
253	ProSerValGlyHisAlaLeuHisGluGluArgVal-----GluArgProPro---	268
76078	CCTTCTCTTCTATCAAAAAGAGAGCATTTGGAGAGTGTACTATGATGAAGTGAACCCCTGAG	76137
269	-----MetPheProProThrPheLeuTyrThrGlnSerTrpGluAspProGluProAsp	286
76138	TACAGCAATTTAAGAACATTTATATATACGATTCACATGGGAAGATCTCTAGAGAAGAT	76197
287	MetGluValMetGluIleAsnProLysAspThrValLeuThrLeuThrSerGlyGlyCys	306
76198	GTTAACATTTCTTAATTAATTAACCAAGAGATACCAATTTGGCTATTACCACTGCTGGTGTAT	76257
307	AsnAlaIleAsn-----LeuLeuValGlnGlyAlaGlyGlnValValSerValAspCys	324
76258	AATATTTTGCACATATGCTACTTTGCTCAATCCACCGAAGAAATTCATGGGGTGGATTTA	76317
325	AsnProAlaGlnSerAlaLeuLeuLeuGluLeuLysValAlaIleGlnGlnLeuGluPhe	344
76318	AATCCATGTCAAGACATTTGACAGAAATTAATAATTAGCAGCATTAGAAGCTTGAGTTTC	76377
345	GluAspValTrpGlnLeuPheGlyGlyValHisProArgIleGluGluLeuTyrGlu	364
76378	ACACAGTTGTGGCAATGTTTGGTGAAGTATGATGATTAACAAATATTTTACTG	76437
365	LysLysAlaLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrpSerLysArgLeuTrp	384
76438	AATAAATTTGGCGCATATCTTTTCATCAACGCATTTTCAATACTGG-----	76482
385	TyrPheGlnHisGlyLeu-----TyrTyrGlnGlyGly	395
76483	---TTCAAAATGAAGACGAAACAAATTTGATCTTAATGGAGCCGGGTGTATGACACTGGG	76539
396	MetGlyLysLeuCysTrpValLeuGlnCys-----LeuAlaValLeuLeuGlyLeu	412
76540	TTCAACAAG-----TGGGCATTAAGACTTGCATAATGGGTATTCAAAAGTTGCAATCTT	76593
413	GlyLysThrValLysArgLeuAlaAsnAlaProThrMetGluGlnGlnArgArgLeuTrp	432
76594	ACAGATGAAGTTAATATGCTATGCAAAAGCTTAGAAGCCCTAGAGGAGCAAGATCTATTGG	76653
433	AspSerAsnMetLeuIleHisPheValLysAsnGlyProLysProLeuValTrpLeuPhe	452
76654	GATAAAAAGATA-----AAGCCGGTC-----CTTTTC	76680
453	ValLysPheValSerLeuValLeuPhe---AsnLysAlaValLeuTrpPheGlyGlyGly	471
76681	AACAGATAGTGGTGAATAATCTCGTTGGAAACCCGTTATTTTATGGAGTCACTTGA	76740
472	ValProGlyLysGlnTyrAlaLeuIleLysAlaAspGlyIleProIleGluAsnTyrIle	491
76741	GTACCACGTAATCA-----GCAAAAATGATGGGCTCTTCAACATTCGAATATATT	76791

```

QY 492 AlaArgThrMetAspGlyValAlaGluAenSerHisValArgLysGlnAenTyrPheTyr 511
DB 76792 ATTGACACTTTGGACCGCTGTTATTGATAATCTTTCGATTTCGATTAATCTTTAT 76851
QY 512 TyrAsnCysLeuThrGlyLysPheLeuArgAspAsnCysProThrTyrTrpLeuArgGluAla 531
DB 76852 TATTTGACATTGAAGGAGAGATATTCGAGTAGAAGTTGCTCGATTACCTTAAAGAGGGT 76911
QY 532 AlaPheAlaThrLeuLys-----SerGlyValValAspAenLeuThrValSer 547
DB 76912 GGATTCAAGTCTCTTCTCGTGAAGCTCGTGAATCTCTCTAGATAGAGTTAGACTTCAT 76971
QY 548 ThrAsnPhePhe-----MetGluGluLeuLysAlaArgThrTyrThrLysValIle 564
DB 76972 ACTGACACGTTGAAGGATGCTCTGAGCGTTCTCGAAGAAACAGTTTCGATTGCTATA 77031
QY 565 LeuMetAspHisValAspTrpLeuAspMetProValAlaAenGluLeuAlaGlu----- 582
DB 77032 ATTATGGATCACATGGATTGTTGAT---CCACAAGGAACCTGATGTTGACGAGAAAT 77088
QY 583 ---CysLeuAlaLysGlnValAlaProGlyGlyIleValIleTrpArgSerAlaSerLeu 601
DB 77089 CAGGCTTTGGTGGTGGCCCTCAACTCTCGCGTAGGTTATTGAGATCTGCTCTAAA 77148
QY 602 SerProProTyrAlaGluLeuIleGlnLysAlaGlyPheAspValArgCysIleArgArg 621
DB 77149 AGCCCATGGTATATCAAAACTTCGAAAGTTCCGGCTTCAGTTGTTAAAGCAGTAAGTGCA 77208
QY 622 AlaThrGlnGly---TyrMetAspArgValAsnMetTyrSerSerPheTyrMetAlaArg 640
DB 77209 AGATACCCCTGGAAATGTATTGATAGAGTTAAACATGATGCCAGTACCTGGGTTTGCAA 77268
QY 641 Arg 641
DB 77269 AAA 77271

```

RESULT 5  
CR382138 20/c

WPCOMMENT

Sequence split into 24 fragments LOCUS CR382138 Accession CR382138

Fragment Name	Begin	End
CR382138_00	1	110000
CR382138_01	100001	210000
CR382138_02	200001	310000
CR382138_03	300001	410000
CR382138_04	400001	510000
CR382138_05	500001	610000
CR382138_06	600001	710000
CR382138_07	700001	810000
CR382138_08	800001	910000
CR382138_09	900001	1010000
CR382138_10	1000001	1110000
CR382138_11	1100001	1210000
CR382138_12	1200001	1310000
CR382138_13	1300001	1410000
CR382138_14	1400001	1510000
CR382138_15	1500001	1610000
CR382138_16	1600001	1710000
CR382138_17	1700001	1810000
CR382138_18	1800001	1910000
CR382138_19	1900001	2010000
CR382138_20	2000001	2110000
CR382138_21	2100001	2210000
CR382138_22	2200001	2310000
CR382138_23	2300001	2336804

Continuation (21 of 24) of CR382138 from base 2000001 (CR382138 Debaryomyces hansenii ch

Alignment Scores:

Pred. No.:	Length:	Matches:
Score:	1,44e-64	110000
Percent Similarity:	864.50	223
Best Local Similarity:	46.1%	105
		Mismatches: 252

Query Match:	25.0%	Indels:	131
DB:	15	Gaps:	20
US-10-620-914-45 (1-648) x CR382138_20 (1-110000)			
QY 41 SerLysLysGlyAspAspHisAlaAlaArgLeuGluSerPheTyrGlyProGlnAlaAla 60			
DB 58985 AGCAAAAGCCCACTGAGCAACACAGTCGTTAGAACCTTTCTATAAAACCAGGCACAT 58926			
QY 61 AlaPhe-----			62
DB 58925 GTGTATGATAAGACTAGAGAGTATTGTTAAAGGACGTAAGAATGCTTGAGGTTGGCC 58866			
QY 63 AlaAlaArgLeuAlaGluArgSerAsnLeuIleTrpValAspLeuGlyGlyThrGly 82			
DB 58865 ACCGACACATCTCTCTAAGAAGAAGATCTTGCTGGGTAGACATTCGAGGGGAACCTGGC 58806			
QY 83 GluAsnValAspMetMetAlaAspTyrIleAspLeuAlaLys---PheLysSerIleTyr 101			
DB 58805 TCCAATATCGAGTACATGCACGAGGTAATGTGATCTCCAGAATTTCAAGGCTGTTAT 58746			
QY 102 ValValAspLeuCysHisSerLeuCysGluValAlaLysLysLysAlaLysAlaLysGly 121			
DB 58745 CTGTTGATTATTCACCATCGTTGTCGAGGTTCGAAAAAAGGTTCGCTGATAGAGA 58686			
QY 122 TrpLysAsnValGlnValValGluAlaAspAlaCysGlnPheAlaProProGluGlyThr 141			
DB 58685 TGGAGCAATGTGATGTTCTTAGTTGTCAGATGCTTGTGATTTCGGTATTGGGTATGAAAG 58626			
QY 142 AlaThrLeuIleThrPheSerTyrSerLeuThrMetIleProPheHisAsnValIle 161			
DB 58625 GCGAATTTGGTGACCTTCTCTACTCATTTGCTATGATTTCCAACTTTTCCATGCTGCT 58566			
QY 162 AspGlnAlaCysSerTyrLeuSerGlnAspGlyLeuValGlyValAlaAspPheTyrVal 181			
DB 58565 GATCATGCTGTCAGTTATTAGACAAAACAGGTGTTATCGCCTGTGTAGACTTCGGTGTG 58506			
QY 182 Ser-----GlyLysTyrAsp-----LeuProLeuArgGln 191			
DB 58505 CAAACCGAGGAGTCTTCGGTGTGTAGAGTGAATACCTTAGGAGGTATTGTTAATAGAAC 58446			
QY 192 MetProTrpSerArgArgPheThrArgSerIlePheAspIleAspIleAspIle 211			
DB 58445 ATCCCATGGGTCTTGAGGAACCTCTCGCGTATTTGGTGTGAAAGCCCAAGGTTTTTTA 58386			
QY 212 GlyProGluArgArgAlaTyrLeuGluGlnLysLeuGluArgValTrpGluGlnAsnThr 231			
DB 58385 GACCTGTGCGAAGACATTTATCTTGAGTACAAATTTGGTACCAATCCCACTCACTGC 58326			
QY 232 GlnGlySerIleProTyrValProTrpLeuArgAlaProTyrTyrValTrpIleGly--- 250			
DB 58325 TACAATAAC-----AAATTGGGACGTTATTCGCTATTTATTTATTTGGGTGGCTGT 58278			
QY 250 -----			250
DB 58277 GACAAATCAAAATCGTCATCTTTTATTAAATAGAATCAATTGTTGGCAACTGAATCCCA 58218			
QY 251 -----ArgLeuPro---SerValGlyHis 257			
DB 58217 TATCTTGTCTCTTCTGATGATGAAGAACAGATATCGATATTCCTATATCTAAGGGTTCAC 58159			
QY 257 -----			257
DB 58157 GAGGTCGCCCTCATGAATTTTCCAAAAAATTTGCTTACCCTATCTATCTACTACCAAGA 58098			
QY 258 -----AlaLeuHisGluArgValGluArgProProMetPheProProThr 273			
DB 58097 GAAATCTGAGAGTATACCTTCGATGAATTAGAATCAATTTTACAGTTTAAAAATCAA 58038			
QY 274 PheLeuTyrThrGlnSerTrpGluAspProGluProAspMetGluValMetGluIleAsn 293			
DB 58037 TATGTGTATGATCTTACTTTGGGAAGATCCTCGTGAGGATCATACATCTTGAAGTTTACT 57978			

Qy	294	ProLysAspThrValLeuThrLeuThrSerGlyGlyCysAsnAlaLeuAsn-----Leu	311
Db	57977	TCAGAGATACCGTTTTGGCTATCACTTCTGCTGGGGCAACATCTTCTCTCAAGCCACT	57918
Qy	312	LeuValGlnGlyAlaGlyGlnValValSerValAspCysAsnProAlaGlnSerAlaLeu	331
Db	57917	TTGCCAAATCCTCCAAGAGAATACATGCGGTGAATTGAATCCATGTCAAACCATCTC	57858
Qy	332	LeuGluLeuLysValAlaIleGlnGlnLeuGluPheGluAspValTrpGlnLeuPhe	351
Db	57857	TTAGAAATGAACTTGCCCTCATTTAGAGCCCTTATCGAAGATCAGATATGCTATT	57798
Qy	352	GlyGluGlyValHisProArgIleGluLeuTyrGluLysLysLeuAlaProPheLeu	371
Db	57797	GGTGAAGGTAATAAATTTAAGGATTTATTAGTTTGACAAATTCAGTCCCATGTC	57738
Qy	372	SerGlnThrSerHisasnPheTrp-----SerLysArgLeuTyrTrpPheGln	387
Db	57737	TCCTCCAACACTTTCCAATATTGGATGGACAGGGGGCAAGACATCAATCCCTAAAGGA	57678
Qy	388	HisGlyLeuTyrTyrGlnGlyGlyMetGlyLysLeuCysTrpValLeuGlnCys-----	405
Db	57677	AAGGGATTGTAT-----GATACAGGTCCACAGGTGGGCATTAAAGTTGCCAAA	57627
Qy	406	---LeuAlaValValLeuGlyLeuGlyLysThrValLysArgLeuAlaAsnAlaProThr	424
Db	57626	TGGGTGTTTTAAGATCTGCGGTATTACCAAGTAGTTTGACATGCTCTGTGAGTGCACA	57567
Qy	425	MetGluGluGlnArgArgLeuTrpAspSerAsnMetLeuIleHisPheValLysAsnGly	444
Db	57566	TTGGAAGAACCAACAGAGGATTTGGAAATGAAAAATC-----	57531
Qy	445	ProLysProLeuValTrpLeuPheValLysPheVal---SerLeuValLeuPheAsnLys	463
Db	57530	---AAGCCAAACC-----TTATTTAACCATTCGTGCTGCTCTGCTTATAGTGACCT	57480
Qy	464	AlaValLeuTrpPheGlyGlyGlyValProGlyLysGlnTyrAlaLeuIleLysAlaAsp	483
Db	57479	ATATTTTATGGAAGCATTTGGGTGTTCTGCTTAACCAAGTGCATGATG-----	57429
Qy	484	GlyIleProIleGluAsnTyrIleAlaArgThrMetAspGlyValAlaGluAsnSerHis	503
Db	57428	GGCAATTTCTATTTTAAATACGTCGTCGATACCTTCGACCCAGTCATTAAAGATCACTT	57369
Qy	504	ValArgLysGlnAsnTyrPheTyrTyrAsnCysLeuThrGlyLysPheLeuArgAspAsn	523
Db	57368	ATATCTGAAGATAACTATTCTTCTACTACTTGACATTTGAAGGTAGTATACCCCCAAAT	57309
Qy	524	CysProThrTyrLeuArgGluAlaAlaPheAlaThrLeu-----	536
Db	57308	TGCCCTGATTATTGACAGAAAAAGGTTATAAAGCCTAACCATCATCAACAGAAAGTCT	57249
Qy	537	LysSerGlyValValAspAsnLeuThrValSerThr-----AsnPheMetGlu	553
Db	57248	AAAGAGGCCCAATTTGCACATATAAGGTATACATACAGACATGCTTTAAACGATGTTT	57189
Qy	554	GluLeuLysAlaArgThrTyrThrLysValIleLeuMetAspHisValAspTrpLeuAsp	573
Db	57188	AGATTGAGCAAGAAATCGTTATCCATTCGCAATCATTTATGGATCATATGACTGGTTCGAT	57129
Qy	574	MetProValAlaAsnGluLeuAlaGluCysLeuAlaLysGln-----ValAlaProGly	591
Db	57128	CCAAACGGTGAAGATGCTTTGAACGAATAATACCTGCTTTGAAAGCTGTTTAAACACCAAT	57069
Qy	592	GlyIleValIleTrpArgSerAlaSerLeuSerProProTyrAlaGluLeuIleGlnLys	611
Db	57068	GGCAGAGCTTATTAGATCTGCTTCTTACAAACCCATGTTACATATAAAACGTTTCGAAGAC	57009
Qy	612	AlaGlyPheAspValArgCysIleArg---ArgAlaThrGlnGlyTyrMetAspArgVal	630
Db	57008	CTCGGCTTCCACTGTGAAGGCTGCTGGTATTAGAGATACCGGTATTTTCGATTGACAGATC	56949
Qy	631	AsnMetTyrSerSerPheTyrMetAlaArgArg	641



```

14924..15111)
/gene="B18P24.030"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAE76261.1"
/db_xref="GI:38566960"
/db_xref="UniProt/TREMBL:Q7SDH4"
/translation="MARSAARPGPGQEAARPTHRFPFPPPAQWKCRCGQQOQTSET
RNRDFAGPLGPNHAPVAVSGTRGDDAVLAKDDAAPQSGIVSGRCPLGVYFVRVPCAGN
ESDMVNEAVRYNPLPIARVKVGLDQMSLVRSPLAVLDSHPCRPPVLAIGFTLRCD
GGGSGRWGTMILVLRVLRRAAALRAALRMSLVSV"
complement(14161..14342)
/gene="B18P24.030"
/number=1
complement(14343..14411)
/gene="B18P24.030"
/number=1
complement(14412..14477)
/gene="B18P24.030"
/number=2
complement(14478..14565)
/gene="B18P24.030"
/number=2
complement(14566..14717)
/gene="B18P24.030"
/number=3
complement(14718..14923)
/gene="B18P24.030"
/number=3
complement(14924..15111)
/gene="B18P24.030"
/number=4
/notes="29 bp poly(g) tandem repeat"
17595..17623
19883..20827
/gene="B18P24.040"
/notes="B18P24.040"
/notes="B18P24.040"
/notes="B18P24.040"
/notes="similarity to HAPB protein, Aspergillus nidulans,
TREMBL:ENHAPB.1
contains EST gb:AA898026, AA898402"
/codon_start=1
/product="related to CCAAT-binding factor HAPB protein"
/protein_id="CAE76262.1"
/db_xref="GI:38566961"
/db_xref="GOA:Q7SGY4"
/db_xref="InterPro:IPR001289"
/db_xref="UniProt/TREMBL:Q7SGY4"
/translation="MMEYAYQPAQNSHNPHMQTYSSAGNSITSPSSOHLPTQTS
PILOTQHQPSPTQGHSMYQTYGVQPNHMGMPLOAAALAAATAAAGTNTYPNPYMA
ADPNLQOSPRMSGVNSKDVRAQPRSPQMNSTPGQRLSVOVSPGVPAQAMLNHA
SRSAVPPMTAAQPPQSPPEMAGVRESPLYNAKQFHLILKRRVAROKLEOLR
LTNKGKPYLHESRHNHARRRPGGGRFLTADQVAERDKVNGDVKGQSGESVVT
ASKTGTGTRKRAESTSGAPNKKAAAPSPEDDDASD"
19883..20827
/gene="B18P24.040"
/number=1
21444..21471
/notes="28 bp gcacgcacgcacgcac tandem repeat"
22226..26083
/gene="B18P24.050"
/join(23226..23780,23851..24177,24239..26083)
/notes="B18P24.050"
/notes="no similarity to N-terminal half of the protein
similarity to S-adenosylmethionine:diacylglycerol
3-amino-3-carboxypropyl transferase btaA, Rhodobacter
sphaeroides, TREMBL:AF329857.1
contains EST gb:BF072578, BF072579"
/codon_start=1
/product="related to S-adenosylmethionine:diacylglycerol
3-amino-3-carboxypropyl transferase btaA"
/protein_id="CAE76263.1"
/db_xref="GI:38566962"

```

```

/db_xref="GOA:Q7SGY5"
/db_xref="InterPro:IPR000051"
/db_xref="UniProt/TREMBL:Q7SGY5"
/translation="MGDNSAMASHGGMGNISLPSLSVTRTLADLNFTNTTTSIIFFT
GVAVLVFLVTTSTNSYRKTKNEDDNEGNPSLSKSLALFCYSCFKPHATAGTCTGTO
QALSPSPYASQADIYDARTGTLKGREDMLAALASOLRYKYRAGLGLGAGGGLSKR
QNGKTCVTVAGTGTGTRRKPIWVDVGGTGWNIEAMKFPVNVSEFKTVLVDFSPS
LCEVARKRFAELGWRNVVICDARKFRLEDYDDEGSGSDSPSLSGMMGETKP
GHAGAEILTMSYSLMMPDFYSIIDSLLELAPHGLIAVDFYASQKVDFTFRNTTG
GLMNHVGFARNFWRSDADRVLSLEPARRDYLYRFGTVLTNARNNTLGAIPYVI
WLGCLKPESTSLPHEIVEHIDALATESPRSSPRLVGKHSSATNALAPAGVCTAPE
MRSKAPNTAIENISANLPLSPFFYQNHWRIRIYDDQLPKHTQFNDEYIYAFWEDSRV
DRELLNLGDDVLAITSAGDNLSTYMSPARVHAIDLNPAQHLLLELUKVASFTIID
YDFWVKIFGEGKHPDFRSLLSKLSPLHSGRAFYWLNSAHIPTDPAGRLGYDTGSSR
YAIRFRFMTSTLFFCRSAVRLLSTPTLEGQSIYHTKPCILLNRFVNLVSSDAF
LWSALGVKPNQVAMTEADYHRRSISSTTSPSKKPSRAEAILHYTTLDPLVLSH
LASDNPYLVCVLGOYTRQCHPDYLSAAHSILSAPGAPDGLRIHTDEIOELARQOP
GTLTVAVYMDSDMDFDPPSPSEKEGKRGKARQVRLNALKVGGKVLRSAGVEPWY
VRVVEEGFGARRVCGRSGRGDQCICIDRVNMTASCIWLEKMEDLEELVDASA"
23226..23780
/gene="B18P24.050"
/number=1
23781..23850
/gene="B18P24.050"
/number=1
23851..24177
/gene="B18P24.050"
/number=2
24178..24238
/gene="B18P24.050"
/number=2
24239..26083
/gene="B18P24.050"
/number=3
27765..28499
/gene="B18P24.060"
complement(join(27766..28118,28258..28275,28370..28499))
/gene="B18P24.060"
complement(join(27766..28118,28258..28275,28370..28499))
/notes="match of 21 ESTs
similarity to succinate dehydrogenase precursor (EC
1.3.99.1), YDR178W, Saccharomyces cerevisiae, PIR:A54380
contains EST gb:AI397618, AW18559"
/codon_start=1
/product="related to succinate dehydrogenase precursor"
/protein_id="CAE76264.1"
/db_xref="GI:38566963"
/db_xref="GOA:Q7SGY6"
/db_xref="InterPro:IPR007992"
/db_xref="UniProt/Swiss-Prot:Q7SGY6"
/translation="NASVARSSALLKQAAQOQVAAGLURVAAFTTTSRKSLLPPPPQ
RLEGTVNDPEVPPSPSHGSHVHTFDRVVAAGLPLTVAPFAAGSLNPTMDAVLAAT
ILIHSTFTGNIIVDYPVSKRVKARKVFTWGLNAAATVLVGLALYEFETDVLGTETI
KRWKA"
complement(27766..28118)
/gene="B18P24.060"
/number=1
complement(28119..28257)

```

## Alignment Scores:

```

Pred. No.:      2,14e-59      Length:      75337
Score:          804.00      Matches:      238
Percent Similarity: 43.2%      Conservative: 112
Best Local Similarity: 29.4%      Mismatches: 256
Query Match:      23.2%      Indels:      205
DB:              15          Gaps:      24

```

US-10-620-914-45 (1-648) x BX842626 (1-75337)

```

Qy      19 LeuGluLeuLeuLeuSerSerMetLysAspAspLeuThrValLeuArgHisMet--- 37
      |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db      23674 TTGGAGGAGCTGGGATG---GGCTTGAANAAGACAAAGGAATGGGACATGTGTA 23730

```

```
QY 38 ---TrpPheGlySerLysLysGlyAspAspHisAlaAlaArgLeuGluSerPheTyrGly 56
Db 23731 CCGTGGCCGGGACAGGACCGGACCAAGGA-----GGAAACCGATATGGG 23775
QY 57 ProGlnAlaAlaAlaPheAlaAlaArgLeuAla-----GluArgSerAsn 71
Db 23776 TAGATGTACGCCCTCCTCATCAGAACCCCTTGCCAATGTTGAAATCCCAACACAGCGGTGACC 23835
QY 72 LeuIleTrpValAspLeuGlyGlyThrGlyGluAsnValAspMetMetAlaAspTyr 91
Db 23836 ATGCTACAAACAGGTGCGTGGGGGACAGCGGTGGAATATCAAGCCATGGCAAGTTT 23895
QY 92 IleAspLeuAlaLys---PheLysSerIleTyrValValAspLeuGlyHisSerLeuGly 110
Db 23896 GTCACAGTCTCTGAATCTCTCAAGACTGTTTACTAGTAGACATTTTCGCCGTCACCTTGT 23955
QY 111 GluValAlaLysLysLysAlaLysAlaLysGlyTrpLysAsnValGlnValAlaGluAla 130
Db 23956 GAAGTGGCTAGGAAGCGGTTTCCAGGCTGGGCTGGGAGAAATGTGAGAGTTATCTGCACG 24015
QY 131 AspAlaCysGlnPheAla-----ProProGluGlyThr----- 141
Db 24016 GATGCTCGCAAGTTTATAGCTTGAGGATATAGAGATGTTGACGAAGGAGAGTCTGCGCTCT 24075
QY 141 ----- 141
Db 24076 GGAGATCTTCGCCCTCTTGTTCGGGTTGGTGGGGGAGACGAAGCGGGACACATGCG 24135
QY 142 ---AlaThrLeuIleThrPheSerTyrSerLeuThrMet----- 153
Db 24136 GGAGCTGAGTTGATCACCAGTCTTATAGCTTTTCGATGATGGTATGTTCTTTGATTAAAG 24195
QY 154 -----Ile-ProProPheHisAsnVa 160
Db 24196 GAGCTCTTGGGGTGGTGTGTGCTGACATATCATGACGATAGTAGCGGATTTATTTCTCGAT 24255
QY 160 IleAspGlnAlaCysSerTyrLeuSerGlnAspGlyLeuValGlyValAlaAspPheTy 180
Db 24256 TATCGATTGCTGAGTCTCTGTGTAGCACCTCAGCGCTTGATTGCGCTGCTGTGACATTTTA 24315
QY 180 rValSerGlyLysTyrAsp-----LeuProLeuArgG1 191
Db 24316 CGCCAGCTCGAAGTCCGACTTCACATTCGCACTACACCGGTGCTTATGAACCGACA 24375
QY 191 nMetProTrpSerArgPhePheTrpArgSerIlePheAspIleAspIleAsp11 211
Db 24376 CGTTGGCTATTTCGCGCGGAACCTTCGCGCTCGTGGTTCGATGCTCAGACGGGTGCTCT 24435
QY 211 eGlyProGluArgArgAlaTyrLeuGluGlnLysLeuGluArgValTrp----- 227
Db 24436 TGAGCCAGCTGCTCGAGATTATCTCGAGTACAGGTTCCGGGACTGCTCGACCGTCAACGC 24495
QY 228 -GluGlnAsnThrGlnGlySerIleProTyrValProTrpLeuArgAlaPro----- 244
Db 24496 CCGCAACAACACACTTGGGAGCAATTCCTACTACTCTGTTGGG-ATGCTCTCAAGAAGC 24554
QY 245 ---TyrTyrVal-----TrpIleGlyAr 251
Db 24555 CCTTTTCTAGCTCGAGTCTACCAACAGAAATTTGGAACACATCGATGCTATTTCGCAGAG 24614
QY 251 qLeuPro----- 253
Db 24615 AGTCCCAAGATCATACCCCGTCTAGTGGGCAACATTTCTTCTCAGACACAATCGGC 24674
QY 254 -----SerValGlyHisAlaLeu-----Hi 260
Db 24675 TAGCTTTTGCAGTCGGCGCACAGCGCGGAGATGCGCTCAAGAGCGCTTCAATACGGCCA 24734
QY 260 sGluGluArgValGluArgPro-ProMet----- 269
Db 24735 TCGAAGACATCTCGGCCAACCTACTCTCTCCGCTCTCTTCTTACCAAAATCACCACCTGGA 24794
```

---

```
QY 270 -----PheProProThrPheLeuTyrT 277
Db 24795 GGATCTACTACGAGGATCAACTCCCGAAGCACACACCCAGTTTCAATGACGAGTACTATCAGC 24854
QY 277 hrGlnSerTrpGluAspProGluProAspMetGluValMetGluIleAsnProLysAspT 297
Db 24855 CCTTTACTTGGGAAGACTCGCGCTCGACAGAGAATCTCTTAACCTCGCGGCCGCGACGAGC 24914
QY 297 hrValLeuThrLeuThrSerGlyGlyCysAsnAlaLeuAsnLeuValGlnGlyAlaG 317
Db 24915 TGTCTCTAGCCATCACACGCGCGGCGACACATTTCTTCTTCTACCTGATGACGAGTCCCG 24974
QY 317 lyGlnValValSerValAspCysAsnProAlaGlnSerAlaLeuLeuGluLeuLysLysV 337
Db 24975 CTGCGGTGACGCGCATCGACCTAAACCCAGCCCAAAACACCACTGCTTGAATCAAAAGTCG 25034
QY 337 alaIleGlnGlnLeuGluPheGluAspValTrpGlnLeuPheGlyGlyGlyValHisP 357
Db 25035 CCTCTTTTACGACTCTCTGGATTATCCCGACGCTCTGGAAGATCTTTCGGTGAGGGCAACACC 25094
QY 357 rArgIleGluGluLeuTyrGluLysLysAlaProPheLeuSerGlnThrSerHisA 377
Db 25095 CCGACTTTCGCTCAGCTCACTCTCCAACTCTCCCTCACCTCTCGCGCGCGCTTTC 25154
QY 377 snPheTrpSerLysArgLeuTrpTyrPheGln-----HisGlyLeuTyrTyrG 393
Db 25155 AATACTGGCTATCCAATGGGCACATATTTACCGACCTCGCGGGCGGCTCTATGATA 25214
QY 393 lnGlyGlyMetGlyLysLeuGlyTrpValLeuGlnCysLeuAlaValValLeuGlyLeuG 413
Db 25215 CCGCGCGCTCCCGATACGCTATCCGTTCTTCGCTGGATTTCCACACTCTTCTTCTGCC 25274
QY 413 lyLysThrValLysArgLeuAlaAsnAlaProThrMetGluGlnGlnArgArgLeuTrpA 433
Db 25275 GCTCCCGGCTCGCTGACCTCTCTCTACTCCACCTCGAAGGGCAACGTTCCATCTACC 25334
QY 433 spSerAsnMetLeuIleHisPheValLysAsnGlyProLysProLeuValTrpLeuPheV 453
Db 25335 ACACCAAGATT-----CGTCCCTGT-----CTGCTCA 25361
QY 453 allyPheValSer---LeuValLeuPheAsnLysAlaValLeuTrpPheGlyGlyGlyV 472
Db 25362 ACGCTTCTGCTCAACGCGCTGCTCAGCTCCGACGCTTCTCTGCTCGGCTTGGGCG 25421
QY 472 alProGlyLysGlnTyrAlaLeuIleLysAlaAsp----- 483
Db 25422 TGCCCAAGATCAAGTGGCTATGTCGAGCCGACTACACCCGCTTCTATCTCTCTCT 25481
QY 484 -----GlyIleProIleGluAsnTyrIleA 492
Db 25482 CCACCAACCCCGACGACAAAGAAAAAACCAAGCGCGCGGAAGCAATCTCCTCACTACAA 25541
QY 492 laArgThrMetAspGlyValAlaGluAsnSerHisValArgLysGlnAsnTyrPheTyrT 512
Db 25542 CTCCACCTCTGATCCCGTCTCTCCACCTCCACCTGCTGCTCGGACACCTTACTTACC 25601
QY 512 yrAsnCysLeuThrGlyLysPheLeuArgAspAsnCysProThrTyrLeuArgGluAlaA 532
Db 25602 TCGCTGTGTCTCTGGGACATACACAGCCAGTCCCATCCGATTACTTTTCCCTGCCG 25661
QY 532 laPheAlaThrLeuLysSer---GlyValValAspAsnLeuThrValSerThrAsnPheP 551
Db 25662 CCCACTCTATACTCAGCGCTCTGAGCCCTTACCGGCTTACGCTTACGACGAGTCAACAGGATA 25721
QY 551 heMetGluGluLeu-----LysAlaArgThrTyrThrLysValIleLeuMetAspH 568
Db 25722 TACAGGAGGTGTGGCTAGGTTTTCAGCGGCTTTCAGCTAGCGGTGATGATA 25781
QY 568 isValAspTrpLeuAspMetPro-----ValAlaA 578
Db 25782 GTATGATGTTGGTTCGATCCCGCTTCGCTCGAGGAGAAAGGAAGGAGGCAAGCGCA 25841
QY 578 snGluLeuAlaGluCysLeuAlaLysGlnValAlaProGlyGlyIleValIleTrpArgS 598
```

```

Db      25842 GGAGGCAAGTGGAGGAGGTTGAATCGCGCGTTGAAGCGTGGTGGAAAGGTGTTTGGAGGA 25901
Qy      598  erAlaSerLeuSerProProTyAlaGluLeuLeuGlnLysAlaGlyPheAspValArgC 618
Db      25902 GC CGCGGAGTGGAGCGCGTGTATGTGAGGGTGTGTTGAGGAGGAGGTTTGGAGCAAGGA 25961
Qy      618 ysile-----ArGArgAlaThrGlnGlyTyrMetAspArgValAsnMetT 633
Db      25962 GCGTGGGTGTCGTGTAATCGGAAGGGGGGACGAGGTGATTGACAGGGTGAATATGT 26021
Qy      633 yrSerSerPheTyrMetAlaArgArg 641
Db      26022 ATCTAGTGTGTCGATCTTCGNAAG 26047

RESULT 7
AJ878596
LOCUS      AJ878596      6962 bp      DNA      linear      PLN 03-JUN-2005
DEFINITION Rhizopus stolonifer abcl gene for putative ABC transporter protein.
ACCESSION  AJ878596
VERSION     AJ878596.2 GI:66947996
KEYWORDS   ABC transporter protein; abcl gene.
SOURCE      Rhizopus stolonifer
ORGANISM   Rhizopus stolonifer
            Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;
            Rhizopus.
REFERENCE   1
AUTHORS    Cernila, B., Cresnar, B. and Breskvar, K.
TITLE      Molecular cloning and characterization of an ATP-binding cassette
            transporter gene from zygomycete Rhizopus nigricans
JOURNAL    Unpublished
REFERENCE   2
AUTHORS    Cernila, B.
TITLE      Direct Submission
JOURNAL    Submitted (28-JAN-2005) Cernila B., Institute of Biochemistry,
            Faculty of Medicine, Univ. of Ljubljana, Vrazov trg 2, SI-1000
            Ljubljana, SLOVENIA
REFERENCE   3 (bases 1 to 6962)
AUTHORS    Cernila, B.
TITLE      Direct Submission
JOURNAL    Submitted (03-JUN-2005) Cernila B., Institute of Biochemistry,
            Faculty of Medicine, Univ. of Ljubljana, Vrazov trg 2, SI-1000
            Ljubljana, SLOVENIA
COMMENT    On Jun 3, 2005 this sequence version replaced gi:59889709.
FEATURES   source
            1..6962
             /organism="Rhizopus stolonifer"
             /mol_type="genomic DNA"
             /db_xref="taxon:4846"
             2758..>6962
             /gene="abcl"
             2758..2763
             /genes="abcl"
             join(2799..4541,4604..5513,5569..>6962)
             /genes="abcl"
             2799..4541
             /genes="abcl"
             /number=1
             2799..2828
             /genes="abcl"
             join(2829..4541,4604..5513,5569..6887)
             /genes="abcl"
             /codon_start=1
             /product="putative ABC transporter protein"
             /protein_id="CAI47726.2"
             /db_xref="GI:66947997"
             /db_xref="GOA:Q5F226"
             /db_xref="InterPro:IPRO01140"
             /db_xref="InterPro:IPRO11527"
             /db_xref="UniProt/TrEMBL:Q5F226"
             /translation="MKRDHSTSRSVTAQSDGDTOLTNTENHDEKILQEIAAABR
            SDVKSSETDKKKKQKLSVPYIKIFRPATKLELLMICIAIFSAIGAMQPISTIIF

```

```

GQFMWTTISTAMASGDYQALVDATHPLVLIPVYMGTAIVLCAYIAQCFWLTGNSQVRR
INLYTHSLRQMSWFDKABESLTRLATQLIQDGISDKFGLVLMVCIQGLAGF
IIAFYKMGRLAVILATIPLLAGTGAAGYFITKYLKSDQDAYEAGSVAEQVFSGR
TVYSPLOMRFAELYSNRLLENAMKTCIRRGQVLGFGGFMFLVLTALSTIDRVPIDP
TREMLTGSDDLVPFAMLIGAMALLQLPNLSAVSGCGAAVKIYSTIDRVPIDP
SEGLKPSQYNANLEPDMFKYPTPDTITILKLNLTIRPGMTVAFPVPSGSGKSTS
VQLIQYDPNAGSVFLDHNLRDYNVNLRSQIGVVSQEPVLFNMSIKONLMLGSK
VQSNBIDVADKCKANCSFVQLPDGQYDMVSHGMLSGGQKQRIALVHOGDLVEQ
LLDDEATSLDQSERLVQAALDAASADRTTIVIAHRLSTIRNADLIVHOGDLVEQ
GSHNELALGVYADLVKQETATKQGVTTPEPDSBELLRERERIEIAOKSAASEL
DEKOTNDHLFVTTGASSVDAYELKERKERKNAKKOSIPMGKVLNOMRSMWLLA
IGVCGNAIAGAVPPCPALLIPARVIALIISPDMSPGPMSTNLYSELFVIGICAFIG
FSSQVISEFVAGERYTKLRSDIFRAFMRQEIQFFDHDNLSGALTSRLAIDSKNVE
LVTKWGDITQIIVTAITGLSTAFSTQWALTILVLCMTPIAFATGESYKHSRFGDK
TKKANEQSGEAGEAIKEIRTVTALNKQNFETKYHRTDHPHLKQKRAYMSIGYA
LQOGITLYTNAVAFIRFMANGMIDFQOFTCMMAIMLTACQVGRASVFTSTLSKA
KYSATAAPDILREPEIDPLEGIBPQINGDIAFENITPRYPARPTDSIENGREN
LHGKSGQTALVPSGCGKSTIGMLQRYDPISTGTVRLDDNNVQKYSNLNLSHML
VQSPILFMTIGENISFGVDSIQVTSQIEDACRASNIHKFIVGLPQGYDTRVGDK
GSQSGQKQRIALVHOGDLVEQVQSLDDEATSLDQSERLVQAALDAASADRTT
TIAHRLSTIRNADLIVHOGDLVEQVQSLDDEATSLDQSERLVQAALDAASADRTT
4542..4603
/gene="abcl"
/number=1
4604..5513
/gene="abcl"
/number=2
5514..5568
/gene="abcl"
/number=2
5569..>6962
/gene="abcl"
/number=3
6887..>6962
/gene="abcl"

intron
exon
intron
exon
3' UTR
ORIGIN

Alignment Scores:
Pred. No.:      1,32e-53      Length:      6962
Score:          724.00      Matches:     201
Percent Similarity: 46.2%      Conservative: 118
Best Local Similarity: 29.1%      Mismatches:  224
Query Match:     15          Indels:      149
DB:              15          Gaps:        21

US-10-620-914-45 (1-648) x AJ878596 (1-6962)

Qy      82  GlyGluAsnValAspMetMetAlaAspTyrIleAspLeuAlaLysPheLysSerIleTyr 101
Db      14  GCCTGGAATATTGAAATGATGAATGATATTTTCCATTAGAACAAATTTGAAAAGTGATT 73
Qy      102 ValValAspLeuCysHisSerLeuCysGluValAlaLysLysLysAlaLysGly 121
Db      74  TTGTTAGACTTGACGCCCTTCACCTTTCCGCAATTCAGGAGAGATTGCTTAAGAGAGGA 133
Qy      122 TrpLysAsnValGlnValValGluAlaAspAlaCysGlnPheAlaPro----- 138
Db      134 TTTAAAAATGTCTCTGCTGCTTGTCAAGAGCGGCTAATTTTCAGATAGGATAGACTA 193
Qy      139 GluGlyThrAlaThrLeuIleThrPheSerTyrSerLeu----- 151
Db      194 GAAGGCGGTGAGGTGTAATCACTGTTCTCTATGC-ATGTAAGTTTAAGAATAGTAGGAT 252
Qy      152 -----ThrMetIleProPro----- 156
Db      253 GTGTATGTGTGTGTATGTATGTGTAAAGTATGCTATTATTAGTCTATATGATGATCAT 312
Qy      157 PheHisAsnValIleAspGlnAlaCysSerTyrLeuSerGlnAspGlyLeuValGlyVal 176
Db      313 TTTATCCAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 372
Qy      177 AlaAspPheTyrValSerGlyTyrAspLeuProLeu----- 189

```

```
Db 373 GTTGATTTTATGCTCTGACGATCTCGTGACGACCGACGACATGGGCTCTCAACAA 432
Qy 190 ---ArgGlnMetProTTPSerArgArgPhePheTTPArgSerIlePheAAspAsn 208
Db 433 AACCGCATGTCGGTGGTTCCTTCGTCACCTTTTGGCAGATGGTTCGAATGGATGCAT 492
Qy 209 IleAspIleGlyProGluArgAlaGlyLeuGluGlnLysLeuGluArgValTTPGlu 228
Db 493 ATCAACCTTCGACCTGCACGTCGAGATATCTCGATACAAATTTGGAACGATCAAGTCA 552
Qy 229 GlnAsnThrGlnGlySer-----IleProTyrValProTTPLeuArgAlaProTyrTyr 246
Db 553 CTCAATCGTGTAAATCATTTTCAATACATACCTTATTG-----ATTCAAGATGCCTTACTAT 606
Qy 247 ValTTPIleGly-----ArgLeuProSerValGlyHisAlaLeuHisGlu 261
Db 607 ATTCGGTCGGTTCCTCGAATGCTAGACAACAGAGATCTCTCGGGATTGGCCATGGGGAG 666
Qy 262 Glu-----GATACAGATAGTGTGACGATCAACGATCATTTGGTGAGCCAAAGTGGAGCATTT 726
Db 667 GATACAGATAGTGTGACGATCAACGATCATTTGGTGAGCCAAAGTGGAGCATTT 726
Qy 263 -----ArgValGluArgProProMetPhePro----- 271
Db 727 AGCTTTCAAACCAAGCAGTCGAGATGGATTATGACCTCTTTTGGCTTGTGCATACACAG 786
Qy 272 ---ProThrPheLeuTyrThrGlnSerTTPGluAspProGluProAspMetGluValMet 290
Db 787 TTTAGATCTTATATTTATGCTTTCACCTTGGGAGATCCTCGTGGGAGTGGAGTATCTG 846
Qy 291 GluIleAsnProLysAspThrValLeuThrLeuThrSerGlyGlyCysAsnAlaLeuAsn 310
Db 847 GATTTGCAAGGAGACACCATGTTTGTCTATCTCTCGCGGTGATATGACCTCGAA 906
Qy 311 LeuLeuValGlnGlyAlaGlyGlnValVal---SerValAspCysAsnProAlaGln--- 328
Db 907 TATGCTTTCAAAGCACAGCCAGGCAATACATTGTGTGATATGAATCCTTGTGAGGTA 966
Qy 329 -----SerAlaLe 331
Db 967 AAATAAATAATTAGAGAGCAGACAATTCGATTTCTTAATATGTTATATAGAACCACTT 1026
Qy 331 uLeuGluLeuLysLysValAlaIleGlnGlnLeuLupPheLysAspValTTPGlnLeuPh 351
Db 1027 ATTGAGCTCAATTCGACGATATCAAGTCTAGATTAACATGACCTTTTGGCGCATGTT 1086
Qy 351 eGlyGluGlyValHisProArgIleGluGluLeuTyrGluLysLysLeuAlaProPheLe 371
Db 1087 TGGTGACGGAATCCATCCGACCTTTTCAATCTTTTACATGMAAAGCTTTTCACTCATCT 1146
Qy 371 uSerGlnThrSerHisAsnPheTTPSerLysArgLeuTyrPheGlnHisGlyLeuTy 391
Db 1147 AAGCAGCATGCTTTCATATATGCTTCAGCAGCTAGATCGTTTTCAGTCGCAAGTTTTA 1206
Qy 391 rTyrGlnGlyMetGlyLysLeuCysTTPValLeuGlnCysLeuAlaValValLeuGl 411
Db 1207 CAAGACGGCTATTTCGGCTGGCTCTCATCTTCGTCGAATGGTGTATTCGTATGCAAG 1266
Qy 411 YLeuGlyLysThrValLysArgLeuAlaAsnAlaProThrMetGluGlnArgArgLe 431
Db 1267 ATTAATGGCGTGGTAGACTCTTGGTGTCTGCGCAACGATTCGAGGAAACAAGACAGAT 1326
Qy 431 uTTPAspSerAsnMetLeuIleHisPheValLysAsnGlyProLysProLeuValTTPLe 451
Db 1327 TTGGGAGGCAAGTTG-----AGACCTGCT-----CT 1353
Qy 451 uPheValLysPheValSerLeuValLeuPheAsnLysAlaValLeuTTP-----PheGl 469
Db 1354 ACTTTCACCATGATTCAAAGATCTTCGAAACCAATGTTTATGTGGAATGCACTTGG 1413
Qy 469 YGlyGlyValPro---GlyLysGln-----TyrAlaLeuIleLysAlaAspGly 484
Db 1414 AGGTAAACAGCCACTAAACAACAACAAGGCTTAATATATATGATGTAGTGCCAAATGAATCAA 1473
```

```
Qy 485 IleProIle-----GluAsnTyrIleAlaArgThrMetAspGly 497
Db 1474 ATGAACATGTTTTTAATGAACATATCTGTCAAGATATATCGAAACACCTTTGGACCT 1533
Qy 498 ValAlaGluAsnSerHisValArgLysGlnAsnTyrPheTyrTyrAsnCysLeuThrGly 517
Db 1534 ATTCCTTCCCATTCACCTCTTTAAAGACGATCAATACCTTTTATCTTTGTTTAAAGCAA 1593
Qy 518 LysPheLeuArgAspAsnCysProThrTyrLeuArgGluAlaAlaPheAlaThrLeuLys 537
Db 1594 CGTTATATGTCATCCAGTTCCTTCTTATCTTACCAATGAAGGTTTCAGTTC-ATTAAA 1652
Qy 538 -----SerGlyVal 540
Db 1653 GGTAAACATATCTGTTTAAATAATGAAAGCTAAATAAAGATGATCAGACATCAGGTGCA 1712
Qy 541 ValAspAsnLeuThrValSerThrAsnPhePheMetGluGluLeuLys-----Ala 557
Db 1713 TTGGATGCTTTCCGTTTACACACTGAATCTATACTGATGTCACCTTCAGGGTATGTCGTCT 1772
Qy 558 ArgThrTyrThrLysValIleLeuMetAspHisValAspTTPLeuAsp----- 573
Db 1773 GATTATCTAATCTGACTGGTGTATGATCATATGATGGATTGGTTGATCCCCAAGATCCT 1832
Qy 574 MetProValAlaAsnGluLeuAlaGluCysLeuAlaLysGlnValAlaProGlyGlyIle 593
Db 1833 AAGAATTAGAAACAGATATTAGAA---ATGAACGCACATTAATTAAGGACAA 1889
Qy 594 ValIleTTPArgSerAlaSerLeuSerProTyrAlaGluLeuIleGlnLysAlaGly 613
Db 1890 GTCTATTGGAGGAGTCAGGAAAGAACCTTGTGATAATCTCTTATTTTGAACATGGA 1949
Qy 614 PheAspValArg-----CysIleArgArgAlaThrGlnGlyTyrMetAspArgValAsn 631
Db 1950 TTCTTGGTGGAACTTTAGCTATACGTCAACAGGTCATCT---ATTGATCGTGTTAAT 2006
Qy 632 MetTyrSerSerPheTyrMetAla 639
Db 2007 ATGTATGCTAGCTTCTATAGGGCT 2030
```

## RESULT 8

AE017345\_05

WPCOMMENT

Sequence split into 15 fragments LOCUS AE017345 Accession AE017345

Fragment Name	Begin	End
AE017345_00	1	110000
AE017345_01	100001	210000
AE017345_02	200001	310000
AE017345_03	300001	410000
AE017345_04	400001	510000
AE017345_05	500001	610000
AE017345_06	600001	710000
AE017345_07	700001	810000
AE017345_08	800001	910000
AE017345_09	900001	1010000
AE017345_10	1000001	1110000
AE017345_11	1100001	1210000
AE017345_12	1200001	1310000
AE017345_13	1300001	1410000
AE017345_14	1400001	1507550

Continuation (6 of 15) of AE017345 from base 500001 (AE017345 Cryptococcus neoformans var)

## Alignment Scores:

Pred. No.:	5.68e-44	Length:	110000
Score:	633.00	Matches:	244
Percent Similarity:	32.8%	Conservative:	103
Best Local Similarity:	23.0%	Mismatches:	222
Query Match:	18.3%	Indels:	493
DB:	15	Gaps:	35

US-10-620-914-45 (1-648) x AE017345\_05 (1-110000)

```
QY 40 GlySerLysGlyAspAspHisAlaAlaArgLeuGluSerPheTyrGlyProGlnAla 59
Db 26888 GGTAGTCAGACCAAGAACCCAC-----CTCAGGCGTTTACGCGGCGCAAGCC 26938
QY 60 AlaAlaPhe----- 62
Db 26939 GATTATATACAGACAGACAGCTCCAGCTCTCAAGGGTAGAGACCATGCTTTCAGCTG 26998
QY 63 ---AlaAlaArgLeu-----AlaGluArgSer 70
Db 26999 TTGGCTGCCCATCTCAAAAGCTCAGCCTATGCTGAGGCTTCCCGTTAGTCACCTCCAAA 27058
QY 71 AsnLeuIleTyrValAspLeuGlyGlyThrGlyGluAsnValAspMetMet----- 88
Db 27059 CCGAGGATATGGGTGGATTGGGAGCGGTACCGGTTGGAAACATTGAAAGATGTACGTC 27118
QY 88 ----- 88
Db 27119 TTTTGGCTCTGTACAGTCAAAACGTTTTTGAACCTGTGACTAATAACCTGAAACCTGTG 27178
QY 89 -----Ala-AspTyrIleAspLeuAlaLysPheLysSerIleTyrValValAspLeuCy 106
Db 27179 CCATAGGAGTAGTATCTTCCATTGACCTACTTGTGCGATATATACATCATCGACCTATG 27238
QY 106 sHisSerLeuCyGluValAlaLysLysLysAlaLysAlaLysGlyTyrLysAsnValGlu 126
Db 27239 TGAACCCCTCTTGAAGTGGCCAGGCGCAAGATCAAGGAGGCGGTGGAAGAATGTCCA 27298
QY 126 nValValGluAlaAspAlaCyGlnPheAlaProGlu----- 139
Db 27299 CGTCTGTGCAAGATGCTAGTAGGTTGTTTGCACAGATGGGAGAGCGGGCAGTGGA 27358
QY 140 -----GlyThrAlaThrLeuIleThrPheSerTyrSerLeuThrMet----- 153
Db 27359 TCCGAGGGAAGTCTGAGCGCGATCACCATGAGCTATTCGCTCTCCATGGCAAGTTTCA 27418
QY 154 -----IleProProPheH 158
Db 27419 GACATTTCCGACCTTTTCTCCGATGAACCTGATTGTTCTCACAAGATACCTCCCTTTT 27478
QY 158 iAsnValIleAspGlnAlaCySerTyrLeu---SerGlnAspGlyLeuValGlyVala 177
Db 27479 ACCAAGTCTTGGATAGATGATCAAGTCTTGATCTCAAGGGGATGATGGCGGTG 27538
QY 177 laAspPheTyrValSerGlyLysTyr-----AspLeuProLeu- 189
Db 27539 TCGACTTTTACACTTCGCGAGAGTAGGTAACAAGGAACGCGTAGCCCTCTCCCTTTT 27598
QY 189 ----- 189
Db 27599 TTTTGACTATCACTATCAGATGGTCACTAATAACTTTCTTTATACATAGGCAATCGGTA 27658
QY 190 -----ArgGlnMetProTyrSerArgArgPhePheTyrPArgSerIlePheAspIleA 207
Db 27659 CAGCGAGCAACGCTGATCTTGTTCTCAAAATGGTCTCGGAGTGTGCTCGACCTCG 27718
QY 207 spAsnIleAspIleGlyProGluArgAlaTyrLeuGluGlnLysLeuGluArgValr 227
Db 27719 ATGCGGTTCATCTTCAGCGGTCTCGAGAGAGTACCTGGGAATACAAAATGGGCACCATCA 27778
QY 227 rpGluGlnAsnThrGlnGlySer-----I 235
Db 27779 AGACGTACATAGCGCGGAACAACATTTTGAACACTTGGTTTCATACAAAATCCCTAGTATG 27838
QY 235 leProTyrValPro---TyrLeuArgAlaProTyrTyrVal----- 247
Db 27839 TCCCTATATAACCTTTTGGCTTGGAGGGGAAAATTTGGATAGATGTGACATATTAATATCAG 27898
QY 248 -----TyrIleGlyArgLeuProSerValGlyHisAlaLeuHisGluGluArgV 264
Db 27899 CTATGCTCTCTCGGT-----GCTCAGCACAACGATGC 27934
QY 264 alGluArgProPro----- 268
```

---

```
Db 27935 TTCTGGCTCCGCAAAATCGTTTACTTTGGAAGCGGTAATCGTCTCGGCCAAAGCGACCT 27994
QY 269 -----MetPheProProThrPheLeuTyrThrGln----- 278
Db 27995 CGGCTCTCCTCACCCCTACCTCGCTTTTACCAACTCTCCATCTGTATTTGGCTCTCCTTC 28054
QY 278 ----- 278
Db 28055 ACCATGCTGGAAATCCAGAAATTAGTTCTAGGACCGAGTGGCGGCACAGCGGACTCAAAC 28114
QY 278 ----- 278
Db 28115 AATTTTGAAGTCGGTGTCTCTTAAGCCGTTTCACTACCATCTGAGAAAGTTCGGTAT 28174
QY 278 ----- 278
Db 28175 TTTTGTGTTGTTCTGTTGATGGAGAAAACTGATGGAGACACAGCGCATGGAGGATACCT 28234
QY 278 ----- 278
Db 28235 TACCTTGAAGAAAAAATCCAGAGCAATTCAGGACTCACATTTACGATGGGTGAGTCGA 28294
QY 278 ----- 278
Db 28295 GCGCCGGGGGGGGGTTTGTGATCCCAATAACACACACCCGAAACTAACTTTTTCGA 28354
QY 279 -----SerTyrGluAspProGluProAspMetGluValMetGluIleA 293
Db 28355 TGTGTTGACTCAATAGACTTGGGAAGATCCGGCGGTGACGTCACAGAAATTCGAACATCA 28414
QY 293 snProLysAspThrValLeuThrLeuThrSerGlyGlyCyAsnAlaLeuAsnLeu----- 311
Db 28415 ACAAAGATGACACACATCTTGCATCCTCTGAGGTGACATGTGCTCCA-TTACGCC 28473
QY 311 ----- 311
Db 28474 TTGACGCCAAACCTCGAGGATTCACGCGCTCGTCACTATTATTCATTTAAATTCAT 28533
QY 312 -----LeuValGlnGlyAlaGlyGln-----ValValSerValAspCysA 325
Db 28534 CAATCTTTACATCATCCAAAACCTCGACCAAGATGAATCTTTACGATGATAGATGA 28593
QY 325 snProAlaGlnSerAlaLeuLeuLeuLysLysValAlaIleGlnGlnLeuGluPheG 345
Db 28594 ACCGCTCAAGGTCAATTTTAGAGTCAAACTTCAGCTATCCAGGATTCGAAATACA 28653
QY 345 luAspValTyrGlnLeuPheGlyGlyValHisProArgIleGluGluLeuTyrGluL 365
Db 28654 ATGATTTCTGGCTCATTTTGGTGAAGCGCGCCACCCAGAGTTTGACAAGCTTTTGACTA 28713
QY 365 yslLysLeuAlaProPheLeuSer----- 372
Db 28714 CCAGCTATCTCCCTCTTGTCTTCGTAAGTCCCGCTCTTAAATAAAAAGCTACAT 28773
QY 373 -----GlnThr-SerHisAsnPheTyrSerLysArg 382
Db 28774 CTTTCTCTAACATGGTTGGGGGTTTACAAGAGACAGCGCTACGCTTATTGGAAGAACCCAT 28833
QY 383 LeuTyrTyrPheGlnHisGlyLeuTyrGlnGlyMetGlyLysLeuCyTyrVal 402
Db 28834 ACGTCTCAGTTCTCAAGCAATTTCTACTTTCGAGGCTACAGCGGC-----TGGGCT 28884
QY 403 LeuGlnCysLeuAlaVal-----ValLeuGlyLeuGlyLysThrValLysArgLeu 419
Db 28885 CTCGCTCGCCAGATTCGATCTTTATCGTGGTGTAGAGGGATGTCAAGCGCTTG 28944
QY 420 AlaAsnAlaProThrMetGluGlnArgLeuTyrAspSerAsn----- 435
Db 28945 TGTGAGCGGACTTCGAGCGGTGAACAGGAAGGATTGGCAAGAAAGATTAGACCGGTT 29004
QY 436 -----MetLeuIleHisPheValLysAsnGlyProLysProLeuValTyr 450
```





```

CDS
complement(3..182)
/locus_tag="RB2"
/codon_start=1
/transl_table=11
/product="hypothetical protein-signal peptide and
transmembrane prediction"
/protein_id="CAD71330.1"
/db_xref="GI:32442812"
/db_xref="UniProt/TREMBL:Q7UZF2"
/translations="MSRLPFSALFAFSVFTVTCGCGGNGNEVIEDTRSPAEOEQED
YEQMBETMDSATNS"
39..260
/locus_tag="RB3"
39..260
/locus_tag="RB3"
/codon_start=1
/transl_table=11
/product="hypothetical protein-signal peptide and
transmembrane prediction"
/protein_id="CAD71331.1"
/db_xref="GI:32442813"
/db_xref="UniProt/TREMBL:Q7UZF1"
/translations="MLFVIFLLHFGWRAGVDFHFVAAASATNSDRNAERKQGEK
QSRHYCSMDARNECEKLVRELLILRISP"
complement(399..1514)
/locus_tag="RB5"
complement(399..1514)
/locus_tag="RB5"
/codon_start=1
/transl_table=11
/product="hypothetical protein-signal peptide and
transmembrane prediction"
/protein_id="CAD71332.1"
/db_xref="GI:32442814"
/db_xref="UniProt/TREMBL:Q7UZF0"
/translations="MTKRLPTQRRLVLAAGLITPLAAASTAKAGFGEWETGSPY
YEDAWYDSEWDGNDPTDEAIGWNETYDRDEALTSSDNDIDWTKNDHGTY
ADYDSDTYDYDHYGSDYDSDNDGNYEYTSYDNDGVDYDAFASYDITDGDG
IYDADYIYFSNAGNDQSKTAQKADQONSAKSQSFYTTIRAKYKVTATNVV
VELQNSQEGSMIADLPADLPQDLNDTLTVGSPFKAGDTSVLLSRSEIHNGQQ
MOVERSGREYSGTIESTKTVDRGQORQAKIKTDGKKMLVDLGRKQDLDTNITGGS
QVTINGPAQVQDRVMLLATKVILDGDTIEISMAQK"
1548..1706
/locus_tag="RB7"
1548..1706
/locus_tag="RB7"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAD71333.1"
/db_xref="GI:32442815"
/db_xref="UniProt/TREMBL:Q7UZE9"
/translations="MBSRSQSVVQETNGRKVDKDIIGAARVAQPNRFGKLSGVAAQ
RGSKQKVQ"
complement(1733..5263)
/locus_tag="RB8"
complement(1733..5263)
/locus_tag="RB8"
/function="DNA synthesis and replication; transcriptions"
/notes="best DB hits: BLAST: pir:D83545; probable helicase
PA0799 [Imported] - Pseudomonas; E=7e-95 embi:CAA67095.1;
(X98455) SNF [Bacillus cereus]; E=8e-94 pir:C72027; swianf
family helicase 2 - Chlamydomophila pneumoniae; E=4e-87 COG:
PA0799; COG0553 Superfamily II DNA/RNA helicases; SNF2
family; E=7e-96 PFAM: PF02803; Thiolase, C-terminal
domain; E=0.11 PF00176; SNF2 and others N-terminal domain;
E=3.9e-64 PF00271; Helicase conserved C-terminal doma;
E=1.2e-24"
/transl_table=11
/product="helicase"
/protein_id="CAD71334.1"
/db_xref="GI:32442816"

```

```

/db_xref="GOA:Q7UZE8"
/db_xref="InterPro:IPR000330"
/db_xref="InterPro:IPR001650"
/db_xref="InterPro:IPR011545"
/db_xref="UniProt/TREMBL:Q7UZE8"
/translations="MDWIAEDQLGSLGLMGAEVRLVDSHDGQFDSRAAMLTVKTPPT
LTQDNGVRFDQVQNRKQQTVALEIVPQVDEDDPGVGVELSSCCGCVAFLRDN
SCPHTLAAAWLQEQIARRNGQLIEFGLGKADAAVAGKQFVDFLLRRKNSRPAP
ADASPTRLQWRVRVNMASQAPLATITPEQKRGKGKGTGRVREYFDLQDNLIT
DPLDSKIASLTVPTGHVDNVAIPEALRMLAGHENVAWQDQNALPVEVFCQLLIT
LQPVVEQSKPKASKSOTDGRASKKKTAARTKSSGKSGKSDDELGNVTATWTM
PSESLDSIAPSDAGLIPDDLIIPDDEKVEVQFQDRLSDVDIGIDVQCEIVVGLSPV
EPAILLADADRIVLCTIADPAAGIIOFLRGDTRDILDHATSRALALAGSDS
LIRVELPQLAGPIVPAPCELVQMRPFAGLKVSRVYESRLRSLIPGDAPELVLP
CLAETGPIRLQDLKAELESSEKIQRFGLSRLSHESHTMEWALTDDHALLGLRHD
AGDESPRIVPEGETFRVGEITPKSLRIKIDDSPWFQMTGTVTVGVDPVLDLLM
AVREDROLVRVDRRFKTISEAFKRLTQGTGLVSEKQLRAGAAVPLARVLDGD
VTELETAHQIAERLSDLSQMSDORPEGLDATLRDQDGYRWLARLSRWGVGGLA
DMGLQKTVTLGILVERGVPALVVAFTSVGENVWREAEKFTPALTKLYRDCDRD
EVIKTAGPNLLIVSYQLQDQAKPASRGWHTLVDEAQFINKSQTKTSQAIRITQA
DWRLGLSGTLENHIGELMSPLTSLPGLSGWQRFSPFABPIERHSDADRESLAR
LVRFVLRRTKDKLVKELPRTETILRAELSAPEKLYEDARLAALAEITAGGAPSH
EGRRRTQTLRWLRLQLACHPSLVEPSWKGTSSKLQFLSLVEELREGDHRALVFSQ
FYKHLVYRAALDERGISYQYLDGATPSHERQRRVDAFQNGEGDLFLISLKAGGTGLN
LTAADYVLLDPWMPAVEDQATRAHRIGQERAVTVYRLVAERTIEEQILQHAOKR
ELVAGILDTDAAKLQTDLLDLKAENG"
complement(5458..5850)
/locus_tag="RB13"
complement(5458..5850)
/locus_tag="RB13"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAD71335.1"
/db_xref="GI:32442817"
/db_xref="UniProt/TREMBL:Q7UZE7"
/translations="MHLPHGNRFKDKTKNYRSETSLTILATVLTVPARHPF
PRSRFELMETPESDPQLSKCKPQPSGTIPGNKVSQSPVIPKBLPHTTCEQPLSNR
AGNRVMDLKRSGVGFGRVPLRLSLP"
5836..7386
/locus_tag="RB14"
5836..7386
/locus_tag="RB14"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAD71336.1"
/db_xref="GI:32442818"
/db_xref="UniProt/TREMBL:Q7UZE6"

```

## Alignment Scores:

Pred. No.:	5,31e-28	Length:	311050
Score:	460.00	Matches:	139
Percent Similarity:	43.4%	Conservative:	96
Best Local Similarity:	25.7%	Mismatches:	210
Query Match:	13.3%	Indels:	96
DB:	1	Gaps:	18

US-10-620-914-45 (1-648) x BX294133 (1-311050)

QY	140	GlyThraAlaThrLeuLeuThrPheSerTyrSerLeuThrMetIleProProPheHisAsn	159
Db	181119	GGTACTGCAACACGGGTG-----ATTGGTGTGAGAACTGCACCGGACTGGTGAAC	181069
QY	160	ValIleAspGlnAlaCysSerTyrLeuSerGlnAspGlyLeuValGlyValAlaAapPhe	179
Db	181068	GTCATACCGGTGACATTCATTGG-----TCGACGCTACGACGAGGACTTATTCCTTC	181015
QY	180	TyrValSerGlyLys-----TyrAspLeuProLeuArgGlnMetPro	193
Db	181014	GTCTGCCACACAGAGAAACGCTGCTGCTGTTGTTTCATGCCACCAACCAAGTA	180955
QY	194	TrpSerArgArgPhePheTrpArgSerIlePheAspIleAspIleGlyPro	213



misc_feature	<p> /db_xref="InterPro:IPR003594"  /db_xref="InterPro:IPR003660"  /db_xref="InterPro:IPR003661"  /db_xref="InterPro:IPR004358"  /db_xref="InterPro:IPR005467"  /db_xref="UniProt/TREMBL:Q6NC38"  /translaton="MQADKQPETHAWRRGLEPKIGLSKLLLTPLILIIABILIYV  PSIANYLRLNRLAANAATAALVLDASPSRVPDSLAQOILDSIARAVALTKMGQV  RLASNLPTKIDHVDNRNVTWGAJLDSFVVMLESQNPALRVVGPASQNAEFISV  IDAAPLNAMYSNLLVSLITSLITGLIYLAHLFLVPRMRLTANMNVFRDRP  ESPASIVVPGRGDELGLARESDMORDLSMHOKSLAALGLAVSKINHDLRLNL  ASSQLSDQSVPDRVQEPAKLVSLERLAFACQSTLSYGRACAPAPDRSILIE  PLVNDRETAGLADASITWFAIERGLTIDADPDQFVLLNLVNAQAALGGQPN  DANIQQIRITREGVSTGATFETSDPGPKARDHLFEAFHGSVVRAGGSLGLATAA  ELVRAHGSENLVREGTIGATFETSDPGPKARDHLFEAFHGSVVRAGGSLGLATAA  order(2944, .3012,3394, .3462)  /locus_tag="RPA0635"  /note="T2 probable transmembrane helices predicted  byTMHMM2.0"  /5394, .4670  /gene="crRNA-Arg1"  /locus_tag="RNA_7"  /5394, .4670  /gene="crRNA-Arg1"  /locus_tag="RNA_7"  /product="crRNA-Arg"  /evidence=not experimental  complement(4784, .5293)  /locus_tag="RPA0636"  complement(4784, .5293)  /locus_tag="RPA0636"  /codon_start=1  /evidence=not experimental  /transl_table=1  /product="hypothetical protein"  /protein_id="CA826080.1"  /db_xref="GI:39647560"  /translaton="MLKNLFLPAANDHEHPALAAALREHNPGLTVSDDLLELCLDELE  PEELQESRLFCFGPQLTQSDVVKYQIGHGAYQFYAPLOYPLAPTAQETEDGQCV  SVLAQSGDAAFPDQIIDITVTLPANIEAPQRDLLAFTRLAEMFWRLSHAIACETMP  AEFRSTRLN"  complement(5320, .6648)  /locus_tag="RPA0637"  complement(5320, .6648)  /locus_tag="RPA0637"  /codon_start=1  /evidence=not experimental  /transl_table=1  /product="hypothetical protein"  /protein_id="CA826081.1"  /db_xref="GI:39647561"  /db_xref="UniProt/TREMBL:Q6NC36"  /translaton="MIPTSPSKATELLQAGSLRGLTSORRLDLWADRVGPVRLQRP  GEAPDRGAGLVI VTEPMAGAGSLFCNALPGDCVVVLPFGENPDFFLSKLTFETV  GPCCAEGPHMGVGVADAGETARPRIVSCYQHGCCDLVMPRLQAAESFGL  AAHIGLMSKEDTDLVLAKEAFILQAWELFPPELILYIAGALLQAPPLLPALQDCV  ALHWNWETLSARTLVFGRSDAKALLGWRLAKAHPAWDGYLLDQAWSLALSRA  LDTVWLPSTHYTLHGDPQAQSAVIVHDLPGTSDADLPDGPASLVTRARRAGTGGED  ALVVWTSHQSGDGVAVILKDITADSRVVATTETATSAFADNNGFGFRLMLCLCA  RDEVDFAREAAASARLVRVAPGQPIPHDFARLDQADQAPSLIARGR"  complement(6725, .9127)  /locus_tag="RPA0638"  complement(6725, .9127)  /locus_tag="RPA0638"  /functions="InterPro IPR001951 COGs COG1011"  /codon_start=1  /evidence=not experimental  /transl_table=1  /product="conserved hypothetical protein"  /protein_id="CA826082.1"  /db_xref="GI:39647562"  /db_xref="InterPro:IPR001951" </p>
gene	
tRNA	
gene	
CDS	
gene	
CDS	



Db 102951 GCATCGGCTTCAGCCGACCGCTCCGGCATCTATGCGGGCTTCCACTGCTATATG--- 103007

Qy 640 ArgArgLysGlyAla 644  
|||||  
Db 103008 CGTCGCTCGGGGCG 103022

RESULT 11  
AF329857  
LOCUS  
DEFINITION Rhodobacter sphaeroides betaine lipid biosynthesis operon, complete  
sequence.  
ACCESSION AF329857  
VERSION AF329857.1 GI:14091029  
KEYWORDS Rhodobacter sphaeroides  
SOURCE Rhodobacter sphaeroides  
ORGANISM Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacteriales;  
Rhodobacteraceae; Rhodobacter.  
1 (bases 1 to 3045)  
Klug, R.M. and Benning, C.  
Two enzymes of diacylglycerol-O'-4'-(N,N,N-trimethyl)homoserine  
biosynthesis are encoded by btaA and btaB in the purple bacterium  
Rhodobacter sphaeroides  
Proc. Natl. Acad. Sci. U.S.A. 98 (10), 5910-5915 (2001)  
1131765

JOURNAL  
PUBMED  
REFERENCE  
AUTHORS Klug, R.M. and Benning, C.  
TITLE Direct Submission  
JOURNAL Submitted (18-DEC-2000) Department of Biochemistry & Molecular  
Biology, Michigan State University, 224 Biochemistry Bldg., East  
Lansing, MI 48824-1319, USA  
Location/Qualifiers  
1. .3045  
/organism="Rhodobacter sphaeroides"  
/mol\_type="genomic DNA"  
/strain="2.4.1"  
/db\_xref="taxon:1063"  
544. .2423  
/gene="betaine lipid biosynthesis operon"  
544. .1794  
/gene="btaA"  
544. .1794  
/gene="btaB"  
/codon\_start=1  
/transl\_table=11  
/product="S-adenosylmethionine:diacylglycerol  
3-amino-3-carboxypropyl transferase"  
/protein\_id="AAK53560.1"  
/db\_xref="GI:14091030"  
/translation="MTQFALTHLPAPPPVARIQGAHVHTSLLSAEGLMERFSLPHG  
LVVQIWEPAVDMALAIAPGDRIVAIASGCVNLSTYLTQSPGSLIADVLSPAHVAL  
GRKLAARTLPDHAFFDLFGRAADLPAGNALLYDRHIAFALDGRSRRYWEARSPPGRR  
IQLFEGFRYHGNALGRFICAAWTLARAACTDLRGFLDCPDIEAQRSPFYVAHIGLPFEA  
PVQALARPALFGLGIPQAQALLAGDGDGDLVLRQRLRLHLCDFPLRENTAF  
LADRRYPFGGALPWPYLEPTAFETLRNAGRVQIENRSLTEALAAPEEIHGFTL  
LQADWMTDQLTALRWQVTAAPGARVIFRTGGAADLLPGRVPEEILGHRADRAA  
GQAGHAADRSALYGGFHLRRRDA"  
1791. .2423  
/gene="btaB"  
1791. .2423  
/gene="btaB"  
/codon\_start=1  
/transl\_table=11  
/product="S-adenosylmethionine:diacylglycerolhomoserine-N-  
methyltransferase"  
/protein\_id="AAK53561.1"  
/db\_xref="GI:14091031"  
/translation="MTDAHAALMDATYRHRRIYDVTRRHFLGDRDLIAELDPGPG  
ARVLEFACCTGFRNLDIRWPCCRLSGLDISOELASARALGRRAATLADGATFE  
ALPLFCTGFRIVLSVALSMIPDMREALREAAHLVPGRLHVVDGQAGLPGAR  
AGLRGWIGRHFHTPRDDLTALGETALGIGGYAEVRSILGGYAILGTLTR"  
2687. .2762  
/product="tRNA-Pro"

gene  
CDS  
tRNA

## ORIGIN

## Alignment Scores:

Pred. No.: 2,04e-23 Length: 3045  
Score: 379.50 Matches: 127  
Percent Similarity: 43.1% Conservative: 68  
Best Local Similarity: 28.1% Mismatches: 182  
Query Match: 11.0% Indels: 75  
DB: 1 Gaps: 20

US-10-620-914-45 (1-648) x AF329857 (1-3045)

Qy 231 ThrGlnGlySerIleProTyrValProTyrLeuArgAlaProTyrTyrValTrpIleGly 250  
|||||  
Db 547 ACGCAGTTCGCTCCACCCACCTGCCC-----GCCCGCCG-----GTGCC 588  
Qy 251 ArgLeuProSerValGlyHisAlaLeuHis-----GluGluArg 263  
|||  
Db 589 CGC-----CAGATCGCGCGCGCGCGCGACGCGACGCTGCTTCTAGCGCGCGAGGACTG 642  
Qy 264 ValGluArgProProMetPheProProThrPhe-----LeuTyrThrGlnSerTrp 280  
|||||  
Db 643 ATCGAGCGG-----ATGTTCTCGCGCTTCTCCACGGCTCTCTATCCGACAGATCG 696  
Qy 281 GluAspProGluProAspMetGluValMetGluLeuAsnProLysAspThrValLeuThr 300  
|||||  
Db 697 GAGGATCCGGCGGTGACATGCGCGCTCCGCCATCCGCCCGCGGACCGGCTGGTGCC 756  
Qy 301 LeuThrSerGlyCysAsnAlaLeuAsnLeuValGlnGlyAlaGlyGlnValVal 320  
|||||  
Db 757 ATCGCCTCGGGCGGTGCAACGTGCTTCTCTATCTCAGCAGGGCGCGGCTCGATCCTC 816  
Qy 321 SerValAspCysAsnProAlaGlnSerAlaLeuLeuGluLeuLysValAlaIleGln 340  
|||||  
Db 817 GCGTGGATCTCTGCGCCCGCCCATGTGGCGCTGGGGCGGCTGAAGCTCGCGCGCGCG 876  
Qy 341 GlnLeu---GluPheGluAspValTrpGlnLeuPheGlyGluGlyValHisProArgIle 359  
|||  
Db 877 ACGTCCCGCGACCATGCGCGCTTCTTCGATCTCTTCGTCGCGCAGACCTCGCGCAAT 936  
Qy 360 GluGluLeuTyrGluLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrp 379  
|||||  
Db 937 GCGGCTCTTACAGCGCCACATCGCGCGCGCGCTCGAGCGCGGCGCGCGCTACTCG 996  
Qy 380 -----SerLysArgLeuTyrTrpPheGlnHisGlyLeuTyrTrpGln 393  
|||  
Db 997 GAGGCGCGCAGCCCTTCGCGCGCGCATCCAGCTGTCGAGCGCGGCTTCTACCGGCAC 1056  
Qy 394 GlyGlyMetGlyLysLeuCysTrpValLeuGlnCysLeuAlaValValLeuGlyLeuGly 413  
|||  
Db 1057 GGTGCTCGCGCGCTTCATCGCGCGCGCGCCATACGCTCGCG-----CGGGCGCGCG 1110  
Qy 414 LysThrValLysArgLeuAlaAsnAlaProThrMetGluGluGlnArgArgLeuTrpAsp 433  
|||||  
Db 1111 ACCGACCTCGGGGCTTCTCGACTCTCCGACATCGAGCGCGCGCGAGCTTCTTCTAC 1170  
Qy 434 SerAsnMetLeuIleHisPheValLysAsnGlyProLysProLeuValTrpLeuPheVal 453  
|||  
Db 1171 GCCCATATC-----GGGCG-----CTCTTCGAG 1194  
Qy 454 LysPheValSerLeuValLeuPheAsnLysAlaValLeuTrpPheGlyGlyValPro 473  
|||  
Db 1195 GCGCGCGTGGTGCAGCGCGCTCGCGCGCGCGCGCGCTCTTCGGGTGGGGATCCCG 1254  
Qy 474 GlyLysGlnTyrAlaLeuIleLysAlaAspGly-----IleProIleGluAsn 489  
|||||  
Db 1255 CCGCGCATATGCGCTTCTGCGGAGACGCGCGACGCGCGCTGCTGCGGGT----- 1308  
Qy 490 TyrIleAlaArgThrMetAspGlyValAlaGluAsnSerHis----- 503  
|||  
Db 1309 -----CTGCGCGAGCGCTCACCGGCTGCTCTGTGACTTC 1344  
Qy 504 ValArgLysGlnAsnTyrPheTyrTyrAsnCysLeuThrGlyLysPheLeuArgAspAsn 523

```

Db 1345 CCCTCGCGGAGAACTACTTCGCTCCAGGCCATCGCCGCGCTATCGCGCGCCGCGC 1404
Qy 524 -----CysProThrTyrLeuArgGluAlaAlaPheAlaThrLeuLysSerGlyVal 540
Db 1405 GAGGGCGCGCTGCGCGCTATCTCGAACCCACCGCTTCGAGACGCTGCGCGAGAAC--- 1461
Qy 541 ValAspAsnLeuThrValSerThrAsnPhePheMetGluGluLeuLysAlaArgThrTyr 560
Db 1462 GCGGGCGGGTGAGATCGAGAACCGGAGCGGCTGACGAGCGCTCGCGCGGACCGGAG 1521
Qy 561 ThrLysVal-----IleLeuMetAspHisValAspTrpLeuAspMetProValAla 577
Db 1522 GAGAGCATCCAGCGCTCACCTGCTCGATGCGCAGGACTGGATGACGGAGCGGAGCTG 1581
Qy 578 AsnGluLeuAlaGluCysLeuAlaLysGlnValAlaProGlyGlyIleValIleTrpArg 597
Db 1582 ACCGCGCTCTGGCGGAGGTGACGCGCACTGCAGCGCGCGCGCGCGGCGGTGATCTCCG 1641
Qy 598 -----SerAlaSerLeuSerProProTyrAlaGluLeuIleGlnLysAlaGlyPhe 614
Db 1642 ACCGGGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1695
Qy 615 AspValArgCysIleArgArgAlaThrGlnGly-----TyrMetAspArgValAsnMet 632
Db 1696 CACTGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1755
Qy 633 TyrSerSerPheTyrMetAlaArgArgLysGlyAla 644
Db 1756 TACGGCGGCTTCCACCTCTACCGCGGCGGAGCGCA 1791

RESULT 12
AR551999
LOCUS AR551999 933 bp DNA linear PAT 08-OCT-2004
DEFINITION Sequence 7130 from patent US 6747137.
ACCESSION AR551999
VERSION AR551999.1 GI:53945174
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 933)
AUTHORS Weinstein,K.G. and Bush,D.
TITLE Nucleic acid sequences relating to Candida albicans for diagnostics
and therapeutics
JOURNAL Patent: US 6747137-A 7130 08-JUN-2004;
Genome Therapeutics Corporation; Waltham, MA
FEATURES
source 1. 933
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Alignment Scores:
Pred. No.: 1.25e-21 Length: 933
Score: 352.00 Matches: 95
Percent Similarity: 40.5% Conservative: 35
Best Local Similarity: 29.6% Mismatches: 97
Query Match: 10.2% Indels: 94
DB: 6 Gaps: 8

US-10-620-914-45 (1-648) x AR551999 (1-933)

Qy 144 LeuIleThrPheSerTyrSerLeuThrMetIleProPheHisAsnValIleAspGln 163
Db 7 TTGATTACTTTTTCGTTATTCATTCGATGATCCAACTTTCAATGCTGCTATCGATAAT 66
Qy 164 AlaCysSerTyrLeuSerGlnAspGlyLeuValGlyValAlaAspPheTyrValSer--- 182
Db 67 GCTGTTTCTAAATAGATATGAGAGGTATATTCCCACTGTAGATTTTGGTATTCAAGC 126
Qy 183 -----GlyLysTyrAsp-----LeuProLeuArgGlnMetPro 193

```

```

Db 127 AGTGACACCTCAATGGGTGCTATTATACTGTTGGGGTGGTTAAACAGGGACATTCCT 186
Qy 194 TrpSerArgArgPhePheTrpArgSerIlePheAspIleAspAsnIleAspIleGlyPro 213
Db 187 TGGATTATTAGTAATTTTGGAGAAATTTGGTTGAAGCTGATAAAGTGTTTGGATTCT 246
Qy 214 GluArgArgAlaTyrLeuGluGlnLysLeuGluArgValTrpGluGlnAsnThrGln--- 232
Db 247 TCAAGAAGAAACTATTTTGGAATATAAATTTGGTACCGTCAAACTCTTTGAATTCATAAC 306
Qy 233 -----GlySerIleProTyrValProTyrLeuArgAlaProTyrTyrValTrpIle 249
Db 307 AAGGCTTTGGTAAATC-----CCCTATTATATTGTAAT 342
Qy 250 GlyArgLeuProSerValGlyHisAlaLeuHisGluGlu-----ArgValGlu 265
Db 343 GGTGTGTAATAATCAAAATCACACCACTTTTGAAGAAGATTGAATGTTTAGCCACTGAA 402
Qy 266 ArgProProMetPheProProThr----- 273
Db 403 TCCCTTACCTTGGTCCAACTCAACTCCAACTCGCTAACTCAACTTGAAGATATTCCAAT 462
Qy 273 ----- 273
Db 463 TCTAAAGTCAATGAAGTCTGCTTTAATCAACTTCCAAAAAATTTACTTACCCATCAATG 522
Qy 273 ----- 273
Db 523 TACTATCAAAAGGAATATTGGAGAGTCTACTATGATGAATGAATCCATTGTATGAACAA 582
Qy 274 -----PheLeuTyrThrGlnSerTrpGluAspProGluProAspMetGluVal 289
Db 583 TTTAAAAAACCAATACATTTATGCTTTCACCTTGGGAAGATCCTCGTGAAGATCATAACTT 642
Qy 290 MetGluIleAsnProLysAspThrValLeuThrLeuThrSerGlyGlyCysAsnAlaLeu 309
Db 643 TTGAATTTTACCAGTGTATGCTGTTTGGCTTACTTACTCAGCTGTGTATATAATTTTG 702
Qy 310 AsnLeu-----LeuValGlnGlyAlaGlyValValSerValAspCysAsnProAla 327
Db 703 AGTTATGCTAGTTTACCAACACCAACCAAAAGATTCATGCTGTGTATCTTAATCCATGT 762
Qy 328 GlnSerAlaLeuLeuGluLeuLysLysValAlaIleGlnGlnLeuGluPheGluAspVal 347
Db 763 CAAAACCATTTATTAGAAATTTGAATTTGGCTAGTTTGTAGATGCTCTTCTCAAGAACAAT 822
Qy 348 TrpGlnLeuPheGlyGluGlyValHisProArgIleGluLeuLeuTyrGluLysLysLeu 367
Db 823 TGCTCAATGTTTGGTGAAGGTAAATTCGAAAAATTTCAATGATCTTTTGATTGATCTTTG 882
Qy 368 AlaProPheLeuSerGlnThrSerHisAsnPheTrpSerLysArgLeuTyrTyrPheGln 387
Db 883 GCGCGGCACATGCTTCTAATGCG-----TTCAA 912
Qy 388 His 388
Db 913 CAC 915

RESULT 13
AE008128
LOCUS AE008128 10129 bp DNA linear BCT 28-MAY-2004
DEFINITION Agrobacterium tumefaciens str. C58 circular chromosome, section 186
of 254 of the complete sequence.
ACCESSION AE008128 AE007869
VERSION AE008128.1 GI:15157252
KEYWORDS
SOURCE Agrobacterium tumefaciens str. C58
ORGANISM Agrobacterium tumefaciens str. C58
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Rhizobiaceae; Agrobacterium.
1 (bases 1 to 10129)
REFERENCE Hankle,G., Slater,S.C. and Goodner,B.
AUTHORS Complete Genome Sequence of Agrobacterium tumefaciens C58
TITLE

```



(Rhizobium radiobacter C58), the Causative Agent of Crown Gall  
Disease in Plants  
Unpublished  
2 (bases 1 to 10129)  
Hinkle, G., Slater, S.C. and Goodner, B.  
Direct Submission  
Submitted (14-AUG-2001) Cereon Genomics, 45 Sidney Street,  
Cambridge, MA 02139, USA  
**FEATURES**  
**source**  
1. .10129  
/organism="Agrobacterium tumefaciens str. C58"  
/mol\_type="genomic DNA"  
/strain="C58"  
/isolate="Cereon"  
/db\_xref="taxon:176299"  
99. .1175  
/gene="AGR\_C\_3839"  
99. .1175  
/gene="AGR\_C\_3839"  
/notes="hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product="AGR\_C\_3839p"  
/protein\_id="AAK87865.1"  
/db\_xref="GI:15157253"  
/translation="MNSGKFGRSALLSSTVGLALASACTSVRYAAKEGNGPKKI  
ATVTSSSAQTSQTPATAVAGAVAGATAPAVQADAAASQAALPTAVAVKGRVAL  
PAMSEIAAAADLATOMQOPEPAGAAATAAPTAAALVPTHTTAAALVETANAAG  
PPAHEAAVDPVAAVKGFPAPPPGPSIGTELTAERKVIPLPKEDSTLVAYAG  
PTVAALAAIQNESNTAPLRTAPAGREKSLGTLTKATYMDVPEDLVHRVVQRESRYN  
AVLSYSGFMGLMOIRATKRSGLGDFGDPASGLFPAETNLKYAVKYLKGAWVVAGNDNRDN  
AGLYAGYGYDAKRGMLVLRN"  
1517. .2767  
/gene="AGR\_C\_3843"  
1517. .2767  
/gene="AGR\_C\_3843"  
/notes="hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product="AGR\_C\_3843p"  
/protein\_id="AAK87866.1"  
/db\_xref="GI:15157254"  
/translation="MTSAAPKTFGSKNTKLKSALLQHKALSGLSERFGVLFGSLV  
YQIWEDEPIDMEALGEGHRIVTIGSGCNMLAYLSRNPASIDVVDNLPHIALNK  
LKLAAPFHLVPAQDVVHFGRGATRSNVCYDRFIAEHLDAATKAYSKRTPLSGRRRI  
SVPDFRNIYRTGLLGRFIGAGHIMARLHGKVLTEMAKTRTLDQORQFDSKVAFLDFK  
VVRWLTKRSSSLFGLGIPPRQYDELASLSDGTVASVLKERLEKACNPLSDNYFAW  
QAPARYEPPEGALPAYLURPEYIEKIRNNTAVVHATYTELLSRKPKANGVDRVIL  
LADQWMTDVLNMLWSQISRTAAAGKARVIFRTAAEKSVIEGRLSPDIRNQWYLYLE  
SNELNADRSATYGGFHLYQRAMA"  
2764. .3483  
/gene="AGR\_C\_3844"  
2764. .3483  
/gene="AGR\_C\_3844"  
/notes="hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product="AGR\_C\_3844p"  
/protein\_id="AAK87867.1"  
/db\_xref="GI:15157255"  
/translation="MKTIGENVGLADSAHAGLMDRMYRQHRIYDITRKYYLIGDRDT  
ISGLDVPGKGTLLIEICGGQGNLLASRRFPDAKFLGLDISAEMLLTASENPAGKAA  
PIILRVADGATTSERISGCGPDGPDVMIPLYALSMIPDEKAIQALALKEGGSLLHVD  
FQCEQLPKWFTLLQAWLTFRHVTFRANLRYVLANNAGRFDGNLVFEELIARGYAWRA  
VITLPVAEAPQPKIHLLADA"  
3434. .4498  
/gene="AGR\_C\_3846"  
3434. .4498  
/gene="AGR\_C\_3846"  
/notes="HYPOTHEICAL 32.0 KDA PROTEIN IN FBAB-THID  
INTERGENIC REGION"  
/codon\_start=1  
/transl\_table=11

```

/db xref="GI:15157260"
/translation="MAROFIYHAGLNKSYGAKKVLNHLSPVDPKIGILGPNAG
KSTVLKIMAGLDEYGEAWLASGATLYLEQEPKLDENKTYMVMWEGVASKTAIVD
RYNELMNYSDETADGAKQDMIDSONLMDQNVEMAMDLKCPFGSDAVTGLSGG
ERRVALCKLLSOPOLLLEDEPNHLDATIAWLEKHLRDYPGAWMIITHRYFLDN
VTGWIILDEGRGIPYEGNYSAYLOKAKEMQOEAREDSRQKAI SREQEWIASSPKA
ROTUKSARIRAYDELVEAENRRPDAQI VIPAERLGRVVI EAENLTKSYGDRVLIE
NLTKLPFGGIVGIPNGAGKTLFRMITGQSPDPSGVTVGETVHLGYVDQSRDIL
AGKTYWEELSGNDLILKLGFEVNSPVCGFNFKGDDOQKVNLSGGORNRLHA
KMLKAGNVLLEDEPNLDLTETLALAEALAPACCAVITSHDRMFLDLRLATHILAF
EGDSHVEWFEQNFEDYKDKVRRLGPESIKPRVSYKRLTR"

```

```

Alignment Scores:
Pred. No.: 6 37e-19 Length: 10129
Score: 336.50 Matches: 180
Percent Similarity: 40.9% Conservative: 107
Best Local Similarity: 25.7% Mismatches: 279
Query Match: 9.7% Indels: 140
DB: 1 Gaps: 26

```

US-10-620-914-45 (1-648) x AE008128 (1-10129)

```

Qy 5 ArgAspGlyArg-----ProAlaSerTyrThrLysLysAsnPheSerLeuGluLysLeu 22
Db 909 CGCGAAGCGCTACAAATCCGGCGCTATAGCGGGCAATTTGGCCCTGATGCAAAATC 968
Qy 23 Lys----- 23
Db 969 CGCGATGCCAGCGCCGCTCCCTGGCTTCGATGGCCCGCCCTCCGGCCTGTCGATGCC 1028
Qy 24 LeuSerSerMetLysAspAspLeuThrValLeuArgHisMetTrpPheGlySerLysLys 43
Db 1029 GAACCAATCTCAATACGCGCTCAATATTTGCGGGCGCATGGTGTGGCGGGTAAC 1088
Qy 44 GlyAspAspHisAlaAlaArgLeuGluSer-----PheTyrGlyProGlnAlaAla 60
Db 1089 GACCGTGACAAATGCCGTCGCTTTATGCGCGGGTATTATTAGCATGC-CAAGCGCAA 1147
Qy 61 AlaPheAlaAlaArgLeuAlaGluArgSerAsnLeuLeuTrpValAspLeuGlyGly 80
Db 1148 GGGCATGCTGCAGCTTCTGGGAA-----CTGACGGCCTCTGCGCTGT 1189
Qy 81 ThrGlyGluAsnValAspMetMetAlaAspTyrIleAspLeuAlaLysPheLysSerIle 100
Db 1190 GCTGGAGACCGCGCGAGCTGA--GCGGCTTTTGAATGTCGCGCTCCCTTCGTGTC 1247
Qy 101 TyrValValAspLeuCysHisSerLeuCysGluValAlaLysLysAlaLysAlaLys 120
Db 1248 GTCAATCTTGGGCTCGTCCCGAGGATCAACATGCGTAAATAATCGATCAGTTGCAGA 1307
Qy 121 GlyTrpLysAsnValGlnValValGluAlaAspAlaCysGlnPheAlaProProGluGly 140
Db 1308 -----ACCTCGGA 1316
Qy 141 ThrAlaThrLeuIleThrPheSerTyrSerLeuThrMetIlePro-ProPheHisAsnVa 160
Db 1317 ATAGCCCGGGATGACAGCTAGTGGTTCAGTGACCACTCGCTTCAGGGCGCGGCC 1376
Qy 160 lIleAspGlnAlaCysSerTyrLeuSerGlnAspGlyLeuValGlyValAlaAspPheTy 180
Db 1377 GTTCTCTGGTGCCGGCTCGGCTGGTGGCGGTGGTTTCCATGTGCTTTCTTTT-- 1434
Qy 180 rValSerGlyLysTyrAspLeuProLeuArgGlnMetProTrpSerArgPhePheTr 200
Db 1435 -----CGCTCTCTGCGACCAACCCCTGTCAACAACTGT----- 1470
Qy 200 pArgSerIlePheAspIleAspAsnIleAspIleGlyProGluArgAlaTyr----- 218
Db 1471 ----AGCGTGGCGCTGTAGAC-----GCCTCTCAATCGACTGAGAGGTGAACATATGACGA 1523
Qy 219 -----LeuGluGlnLysLeuGlu----ArgValTrpGluGlnAsnTh 231

```

```

Db 1524 GTCGGCGACCCAGAGACGGCTTCAGCAAAACACGAAATGAAAGTCGCGATTCGTCACG 1583
Qy 231 rGlnGlySerIleProTyrValProTrpLeuArgAlaProTyrTyrValTrpIleGlyAr 251
Db 1584 ACAAGGCACTCTCCAAAGCGGCT----- 1608
Qy 251 gLeuProSerValGlyHisAlaLeuHisGluGluArgValGluArgProProMetPhePr 271
Db 1609 -----GTCCGAACGGTTTTTTCGGCGCTCTCTTTTCCG 1640
Qy 271 oPro-ThrPheLeuTyrThrGlnSerTrpGluAspProGluProAspMetGluValMetG 291
Db 1641 GCCTCGTC-----TATCCGCAATCTGGGAAGACCCCGAGATCGCATGAGACGGATGG 1694
Qy 291 luIleAsnProLysAspThrValLeuThrLeuThrSerGlyGlyCysAsnAlaLeuAsnL 311
Db 1695 AGCTTGGGAGGCCACCGCATCGTCACCATCGCTCCGGCTCCGGCGGTGCAACATGCTGGCCT 1754
Qy 311 euLeuValGlnGlyAlaGlyGlnValValSerValAspCysAsnProAlaGlnSerAlaL 331
Db 1755 ATCTCTCGGCAACCCCGCCAGCATCGATGTGTGGACCTCAACCCGCCACCATCGCGC 1814
Qy 331 euLeuGluLysLysValAlaIleGlnLeu----GluPheGluAspValTrpGlnL 350
Db 1815 TGAACAAGCTGAAGCTCGCTGCTCCGCCATCTGCCGCCATCAGGATGTGGTGGCC 1874
Qy 350 euPheGlyGluGlyValHisProArgIleGluGluLeu-TyrGluLysLysLeuAlaPro 369
Db 1875 ACTTCGCGCGCGCGG-CACCCGCAACAGCGTGGTTATGACCGTTTCATCGCGCAG 1933
Qy 370 PheLeuSerGlnThrSerHisAsnPheTrpSerLys-----Arg 382
Db 1934 CATCTGGATGCCAGCAAGGCATACTGTGCAAGCGCACCTTTCCGCGCGCGCTGCGC 1993
Qy 383 LeuTrpTrpPheGlnHisGlyLeuTyrTyrGlnGlyMetGlyLysLysCysTrpVal 402
Db 1994 ATTTGGTGTTCACAGCAATCTACCGGACCGGCTGCTGCGCGCTTTTCATCGCGC 2053
Qy 403 LeuGlnCysLeuAlaValValLeuGlyLysThrValLysArgLeuAlaAsnAla 422
Db 2054 GGCCACATCATGCGCGCTGCGCGGTGCAAACTACC-----GAAATGGCCCAAGC 2107
Qy 423 ProMetMetGluGlnArgArgLeuTrpAspSerAsnMetLeuIleHisPheValLys 442
Db 2108 CGGACGCTGACGACGACGCGCAGTTTGTGACAGCAAGTTCGCGCGCTTTTC- 2161
Qy 443 AsnGlyProLysProLeuVal----TrpLeuPheValLysPheValSerLeuValLeuPhe 461
Db 2162 -----GACAAAGCGGTGGTGGCTGGCTGACGAAGCGCAAGAGCTCGCTT- 2206
Qy 462 AsnLysAlaValLeuTrpPheGlyGlyValProGlyLysGlnTyr-----Ala 478
Db 2207 -----TTGCGCTTGGATTTCCGCGCGCCGCGCATGATGACGAGCTGGCA 2248
Qy 479 LeuIleLysAlaAspGlyIleProIleGluAsnTyrIleAlaArgThrMetAspGlyVal 498
Db 2249 AGCTTTTCCAGCAGCGGC---ACGGTTGCTCTCGCTCAAGGAGCGCTGGAAAAGCTT 2305
Qy 499 AlaGluAsnSerHisValArgLysGlnAsnTyrPheTyrTyrAsnCysLeuThrGlyLys 518
Db 2306 GCCTGCAACTTCCGCTC---AGCGCAATATTATTCGCTGGCAGCGCTTTGCGCGCGCT 2362
Qy 519 PheLeuArgAspAsn-----CysProThrTyrLeuArgGluAlaAlaPheAlaThr 535
Db 2363 TATCCGAGCGCATGAGGGTGGCTCGCTTATCTCAAGCGCGCATATACGAAAG 2422
Qy 536 LeuLysSerGlyValValAspAsnLeuThrValSerThrAsnPheMetGluGluLeu 555
Db 2423 ATCGCAACAACCGCG---CGGTCGCGGTGCATCAGCCACCTATACGAGCTGCTT 2479
Qy 556 Lys-----AlaArgThrTyrThrLysValIleLeuMetAspHisValAspTrpLeu 572
Db 2480 TCCGCAAGCGCGCAATATGCGTGCATATATCTGCTCGATGCGCAGGACTGATG 2539

```

Qy 573 AspMetProValAlaAenGluLeuAlaGluCysLeuAlaLysGlnValAlaProGlyGly 592  
 Db 2540 ACCGATGTCAGCTCAACGAGTTATGTCGCAGATCAGCGCACTCGCGGCGCA 2599  
 Qy 593 IleValIleTrpArgSerAlaSerLeu-----SerProTrpTrpAlaGlu 607  
 Db 2600 CGGTCATCTCCGCCACCGCGCGCA-AAAGAGGGTTATCGAGGCGCGGTTTCGCCCGCA 2658  
 Qy 608 LeuIleGlnLysAlaGlyPheAspValArgCysIleArgAlaThrGlnGlyTyrMet 627  
 Db 2659 CATCCGCAACAGTGGGTCTCTCTCGAAGAGCGCTCCACGA---ACTCAAC--GCCATG 2713  
 Qy 628 AspArgValAenMetTyrSerSerPheTyrMetAlaArgLysGlyAlaLysLys 646  
 Db 2714 GACCGCTCGGCCATTATGCGGCTTCATATCACCAGAGGCTATGGCATGAAA 2770

RESULT 14  
 LOCUS AE009162 10134 bp DNA linear BCT 28-MAY-2004  
 DEFINITION Agrobacterium tumefaciens str. C58 circular chromosome, section 188  
 of 256 of the complete sequence.  
 ACCESSION AE009162 AE008688  
 VERSION AE009162.1 GI:17740578  
 KEYWORDS Agrobacterium tumefaciens str. C58  
 SOURCE Agrobacterium tumefaciens str. C58  
 ORGANISM Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 Rhizobiaceae; Agrobacterium.  
 REFERENCE 1 (bases 1 to 10134)  
 Wood, D.W., Setubal, J.C., Kaul, R., Monke, D., Chen, L., Wood, G.E.,  
 Chen, Y., Woo, L., Kitajima, J.P., Okura, V.K., Almeida, N.F., Jr.,  
 Zhou, Y., Bovee, D., Grant, C., Chapman, P., Clendenning, J., Deatherage, G.,  
 Gillet, W., Grant, C., Guenther, D., Kutyavin, T., Levy, R., Li, M.,  
 McClelland, E., Palmieri, A., Raymond, C., Rouse, G.,  
 Saenphimmachak, C., Wu, Z., Gordon, D., Eissen, J.A., Paulsen, I.,  
 Karp, P., Romero, P., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M.,  
 Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C.,  
 Zhao, Z., Dolan, M., Tingey, S.V., Tomb, J., Gordon, M.P., Olson, M.V.  
 and Nester, E.W.  
 The genome of the natural genetic engineer Agrobacterium  
 tumefaciens C58  
 Science 294 (5550), 2317-2323 (2001)  
 JOURNAL 11743193  
 PUBMED 2 (bases 1 to 10134)  
 REFERENCE Wood, D.W., Setubal, J.C., Kaul, R., Monke, D., Chen, L., Wood, G.E.,  
 Chen, Y., Woo, L., Kitajima, J.P., Okura, V.K., Almeida, N.F., Jr.,  
 Zhou, Y., Bovee, D., Grant, C., Chapman, P., Clendenning, J., Deatherage, G.,  
 Gillet, W., Grant, C., Guenther, D., Kutyavin, T., Levy, R., Li, M.,  
 McClelland, E., Palmieri, A., Raymond, C., Rouse, G.,  
 Saenphimmachak, C., Wu, Z., Gordon, D., Eissen, J.A., Paulsen, I.,  
 Karp, P., Romero, P., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M.,  
 Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C.,  
 Zhao, Z., Dolan, M., Tingey, S.V., Tomb, J., Gordon, M.P., Olson, M.V.  
 and Nester, E.W.  
 Direct Submission  
 Submitted (27-SEP-2001) Department of Microbiology, University of  
 Washington, 1959 NE Pacific Ave, Box 357242, Seattle, WA  
 98195-7242, USA  
 FEATURES  
 source  
 1. 10134  
 /organism="Agrobacterium tumefaciens str. C58"  
 /mol\_type="genomic DNA"  
 /strain="C58"  
 /isolate="U. Washington"  
 /db\_xref="taxon:176299"  
 99. 1175  
 /gene="Atu2117"  
 99. 1175  
 /gene="Atu2117"  
 /note="identified by sequence similarity; putative; ORF  
 located using Blastx/Glimmer"  
 /codon\_start=1

/transl\_table=11  
 /product="lytic murein transglycosylase"  
 /protein\_id="AAL43107.1"  
 /db\_xref="GI:17740579"  
 /translation="MRNSGKRGSRALLSSSTVGLALALACTSVEVAKEGPNPKI  
 AVTSSSQAQTSQPAARAVAGATPAAPAVOTADAAASSONAALPTAVAVKGRVAL  
 PPAEEIAAANAIAATQOQTEVAGAAATTAFTTALAALVTHHTTPAAAATETANAAR  
 PAMNAADVLQVAAVKGSPFAAPPSPGSIETELTAERKVIPLPKPDSVLVLAAG  
 PTAAALAIQNESMTAPARTAPAGREKLSGLITKYATWYDVEDLVHRRVQRESRYN  
 PGAYSGNQLQMRHATARSLSGLGDPASGLFDAETNLKYAVKYLGAWVAVAGNDRON  
 AVLYARGVYDAKRKGMHLHLRN"  
 complement(1274..1516)  
 /gene="Atu2118"  
 complement(1274..1516)  
 /gene="Atu2118"  
 /note="putative; ORF located using Glimmer"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAL43108.1"  
 /db\_xref="GI:17740580"  
 /translation="MFTSOSIERLQPPRYSLMTGVAGRESEKKRPLENQAPAPARAGH  
 ORTAAPLEGESVVTDSHVITPGAIVLQIDFLRHV"  
 1517. 2767  
 /gene="Atu2119"  
 1517. 2767  
 /gene="Atu2119"  
 /note="identified by sequence similarity; putative; ORF  
 located using Blastx/Glimmer"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AAL43109.1"  
 /db\_xref="GI:17740581"  
 /translation="MTSAPKTCFSKNTKLKALLQKALSLSGLSERFGLVPSGLV  
 YQIWDPEIDHEAMELGHRIVTIGSCGNMAYLSRNPASIDVVDLNPHEIALNK  
 LKLAAPHPAHQDVVRHFRAGTRNSGYDRIAEDLDAITKAYNSKTLGSRRI  
 SVFDRNYTIGLLFRIGAHIMARLHVGLKTEMAKTRTLDQEQFDSKVAFDFRP  
 VWRWLTFRKSSILGIGIPRQYDELASLDGTVASVLERLEKLAACNPLPSNDYFAP  
 QAFARYPPHEGALPAYLKPEYKIRNNTARVAVHATYTELLSRKPANGVDRIYL  
 LDAQDMTDVQLNEGLSJSRTAASGARVIFRTAAEKSVIEGRSLSDIRNQWYLEER  
 SNELNMDSALYGGFHIYQRAMA"  
 2764. 3483  
 /gene="Atu2120"  
 2764. 3483  
 /gene="Atu2120"  
 /note="identified by sequence similarity; putative; ORF  
 located using Blastx/Glimmer"  
 /codon\_start=1  
 /transl\_table=11  
 /product="O-methyltransferase"  
 /protein\_id="AAL43110.1"  
 /db\_xref="GI:17740582"  
 /translation="MKTIGENVGLADSAHAGLMDRMYRQRIHYDITRKYLLGRDRT  
 ISGLDVPGKGTLLGTCGTGRNLLASRFPDPAKFGLDISAEMLLTASENFAGKER  
 PILRVADATAFRSSEFGPDGFRVMIPYALSMIPDWEKAEQALAAKPGGSIHIVD  
 FQOEOLPKWFTLLQWLTRFHVTFRANLRYVLANAGRPDGNLVFEETARGVAMWA  
 VITLPVAEAPQPKIHRLLADA"  
 3447. 4498  
 /gene="lyc"  
 /note="synonym: Atu2121"  
 3647. 4498  
 /gene="lyc"  
 /note="identified by sequence similarity; putative; ORF  
 located using Blastx/Glimmer"  
 /codon\_start=1  
 /transl\_table=11  
 /product="glycosyl hydrolase/lysozyme"  
 /protein\_id="AAL43111.1"  
 /db\_xref="GI:17740583"  
 /translation="MRKSGDTRARRFAAGAWATIIIRVPMRLRFSALLSALLVSGCTS  
 TSYDLLETASVSRPKFSDPDQDFGVNPHRHEHIGDYSKNGNDWQTVRKSGVSF  
 VFIKATEGSDRIDPKFGDHWRSAAANILHAPVHYFYFCSTADQADWFISNPKVAE

```

TLPPVLVDVENVSPSTCKTRPAPGIVRAEMQRFLEAHYGRKPIIYTSVDFHRDNL
VGQFDYHFWRSVAAPAKIYEDRKWAFQWYATATGWPGVNGPTDINVFAGSEKNR
KWTASAK"
/notes="synonym: Atu2122"
/notes="mitB"
/notes="synonym: Atu2122"
/notes="mitB"
/notes="synonym: Atu2122"
/notes="mitB"
/notes="synonym: Atu2122"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="lytic murein transglycosylase"
/protein_id="AAL43112.1"
/db_xref="GI:17740584"
/translations="MTFRSRISVFAAALSLISAGSALATPPVASDDPKQVACGGDL
GAFLEGKAEAVGAPFAAEIKALAGAQIDPKVLSDRAQGVFRQTLEFFSQTVSQ
ARLDIGKRLQELSSVFAAEIKALAGAQIDPKVLSDRAQGVFRQTLEFFSQTVSQ
HDCRRPELFPOLLALIEMVHGDLDPATNTGAWAGEIGQVOMLPKDIITAFVGDGDN
GHNVDKSPADALITAAKFTQHLGPKKGPWQEVSPENLPWEKLGSLTAGDWF
ALGVTDRGNKPNASIPALIMPQGRKPTFTFYFNFIYLEWQSPFIVTTSAAVFN
RLMGAPVYKNGNDPDPDDATMKELQTLKLEARGHDVGVKVDGILGSGTRVAIQEQRRL
GMPADGWPTQALLQAL"
complement (6313. .7203)
/notes="metF"
/notes="synonym: Atu2123"
complement (6313. .7203)
/notes="metF"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="5,10-methyltetrahydrofolate reductase"
/protein_id="AAL43113.1"
/db_xref="GI:17740585"
/translations="MRSETARRDGDRLSPFPPPKNAEMDQALWETVAELKKWDPDF
VSVYAGAGSTKAPTLDVRRMIDGSHLPTAAHLTCVDASRDVHOVIEFRNAGVRH
FVALRGDSGTICTAQHPPEGVENAAVLKGLREIGDPEISVSAYPEKHPEASSTV
DIDMLKADGNATRALTPFDNDVFERWERTVTAAGITIPVPGIMPLQNLTKLR
FAGRCGTSVPRFIDRPEGEDELEGRARVAEVAEADLQRRRGNFPHLYTMNRA
PLVDVLENLGLQLRRSA"
complement (7208. .8221)
/notes="Atu2124"
complement (7208. .8221)
/notes="Atu2124"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, ArsR family"
/protein_id="AAL43114.1"
/db_xref="GI:17740586"
/translations="MAFGDLKLVLLKAAAGEPTFRFRILLALLSAGDLTVDLTILGQS
OPRISHLLKLTREGIERVOEGAVFRLRQDGEPAVVVRQLIGRASPDAVLORDS
ERLATVKVPAERQAQYFSNANADELRELVSDRAVEGALLKLVGTPIDSLDLG
TGRIQLLSGIYRAIGVDSRDLMSVARNLDRKAGVNASVRGIDNLNPLEQGS
FDLIVHQVLHFDQPEALUSEAARMRLPGGRILVIDLAPHSPFYLDRDEHAHVLGFS
HPLMEWMKAGLSLERAVDLHSEKASDEALDVTIIVWARDQRLIMADDDVVQEAFLRS
GRA"
complement (8383. .10032)
/notes="chvD"
/notes="synonym: Atu2125"
complement (8383. .10032)
/notes="chvD"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="ABC transporter, nucleotide binding/ATPase
protein"
/protein_id="AAL43115.1"
/db_xref="GI:17740587"
/translations="MAFQFIYHAGLNKSYGAKKVLNVHLSFFPDAKIGILPNGAG

```

## Alignment Scores:

```

Pred. No.: 6,38e-19 Length: 10134
Score: 336.50 Matches: 180
Percent Similarity: 40.9% Conservative: 107
Best Local Similarity: 25.7% Mismatches: 279
Query Match: 9.7% Indels: 140
DB: 1 Gaps: 26

US-10-620-914-45 (1-648) x AE009162 (1-10134)

QY 5 ArgAspGlyArg-----ProAlaSerTyrThrLysLysAsnPheSerLeuGluLysLeu 22
Db 909 CGCGAAGCGGCTACAAATCCGCGGCTATAGCGCGCAATTTTCGGCTGATGCAATC 968

QY 23 Lys----- 23
Db 969 CGCGATCCGACGCGCGCTCCCTCGGCTTCGATGGCCCGCCGCTCCGCGCTTCGATGCC 1028

QY 24 LeuSerSerMetLysAspLeuThrValLeuArgHisMetTrpPheGlySerLysLys 43
Db 1029 GAAACCAATCTCAATACCGCGCTCAATATTTCGGCGCGCATGGTGTGGCGGGTAAC 1088

QY 44 GlyAspAspHisAlaAlaArgLeuGluSer-----PheTyrGlyProGlnAlaAla 60
Db 1089 GACCGTGACAAATGCGCTCGGCTTATGCGCGGGTTATTATTACGATGC-CAAGCGCAA 1147

QY 61 AlaPheAlaAlaArgLeuAlaGluArgSerAsnLeuIleTrpValAspLeuGlyGly 80
Db 1148 GGCATGCTGCGACGTTCGCGAA-----CTGACGGCTCTGCCTGT 1189

QY 81 ThrGlyGluAsnValAspMetMetAlaAspTyrIleAspLeuAlaLysPheLysSer 100
Db 1190 GCTGGAGACGCGCGGAGCTGA--GGGCTTTTTCGAAATGCTGGCTCCCTTCGTGTC 1247

QY 101 TyrValValAspLeuCysHisSerLeuGluValAlaLysLysLysAlaLysAla 120
Db 1248 GTCATTTCTGGGCTCGTCCGAGGATCAACATCGCGTAAAAAATCGATCAGTTGCGA 1307

QY 121 GlyTrpLysAsnValGlnValValGluAlaAspAlaCysGlnPheAlaProGlu 140
Db 1308 -----ACCTCGGA 1316

QY 141 ThrAlaThrLeuIleThrPheSerTyrSerLeuThrMetIlePro-ProPheHis 160
Db 1317 ATAGCCCCGGGATGAGCTAGTGTAGTACACACCTCGCTCCAGCGGCGCGCC 1376

QY 160 lIleAspGlnAlaCysSerTyrLeuSerGlnAspGlyLeuValGlyValAlaAsp 180
Db 1377 GTTCTCTGGTCCGCGCTCGGCTGGTGGCGGTGTTTCCCAATGCTCTTTCTTTT 1434

QY 180 rValSerGlyTyrAspLeuProLeuArgGlnMetProTrpSerArgArgPheTr 200
Db 1435 -----CGCTCTCTCTGCGACACCCCTGTCAAACTGT----- 1470

QY 200 pArgSerIlePheAspIleAspAsnIleAspIleGlyProGluArgArgAlaTyr 218
Db 1471 ----AGCGTGGCGGTGTAGAC---GCCTCTCAATCGACTGAGAGGTGAACATATG 1523

QY 219 -----LeuGluGlnLysLeuGlu---ArgValTrpGluGlnAsnTh 231
Db 1524 GTGGCGCACCCCAAGACCGGCTTCAGCAAAACAGAACTGAAGTCGCGCTGCTCCAGC 1583

QY 231 rGlnGlySerIleProTyrValProTrpLeuArgAlaProTyrTyrValTrpIleGly 251
Db 1584 ACAAGGCACTCTCCAAAAGCGGCT----- 1608

QY 251 gLeuProSerValGlyHisAlaLeuHisGluArgValGluArgProProMetPhe 271
Db 1609 -----GTCCGACCGGTTTTTTCGGCGTCTCTTTTCCG 1640

QY 271 oPro-ThrPheLeuTyrThrGlnSerTrpGluAspProGluProAspMetGluVal 291
Db 1609 -----GTCCGACCGGTTTTTTCGGCGTCTCTTTTCCG 1640

```

Db 1641 GCCTCGTC-----TATCCGAGATCTGGGAAGACCCGAGATCGACATGCAAGCGATGG 1694  
Qy 291 luleAsnProLysAspThrValLeuThrLeuThrSerGlyGlyCysAsnAlaLeuAsnL 311  
Db 1695 AGCTTGGCGAAGGCCACCGCATCGTCAACATCGCTCCGGCGCTGCAACATCGTGGCT 1754  
Qy 311 euLeuValGlnGlyAlaGlyGlnValSerValAspCysAsnProAlaGlnSerAlaL 331  
Db 1755 ATCTCTCGCCCAACCCGCGCAGCATCGATGGTGGACCTCAACCCGACCATCGCGC 1814  
Qy 331 euLeuGluLeuLysValAlaLeuGlnLeu---GluPheGluAspValTrpGlnL 350  
Db 1815 TGAACAGCTGAAGCTCGCTCGCTCCGCCATCTGCCGCCCATCAGATGTGTGGCGC 1874  
Qy 350 euPheGlyGluGlyValHisProArgIleGluGluLeu-TyrGluLysLysLeuAlaPro 369  
Db 1875 ACTTCGGCGCGCGCGG-CACCCGCGCAGCAACAGCGTGGTATGACCGTTTCATCGCGGAG 1933  
Qy 370 PheLeuSerGlnThrSerHisAsnPheTrpSerLys-----Arg 382  
Db 1934 CATCTGGATGCCAGCAAGGCATATCTGGTGAAGCGCACCTTTCCGGCCCGCGTCGC 1993  
Qy 383 LeuTrpTyrPheGlnHisGlyLeuTyrTyrGlnGlyMetGlyLysLeuCysTrpVal 402  
Db 1994 ATTTCGGTGTTCACAGGAACATCTACCGACCGCTCGCTCGCGCTTTTCATCGCGGCC 2053  
Qy 403 LeuGlnCysLeuAlaValValLeuGlyLeuGlyThrValLysArgLeuAlaAsnAla 422  
Db 2054 GGCCACATCATGCGCGCTGACGGCGTGAACCTCACC-----GAATGGCAAGACC 2107  
Qy 423 ProThrMetGluGlnArgArgLeuTrpAspSerAsnMetLeuLeuHisPheValLys 442  
Db 2108 CGACCGCTGGACCAACAGCCCGAGTTTTCGACAGCAAGGTCCGCGCTTTTC----- 2161  
Qy 443 AsnGlyProLysProLeuVal---TrpLeuPheValLysPheValSerLeuValLeuPhe 461  
Db 2162 -----GACAGCGCGTGTGTGGCTGGTGCAGAGCGCAAGAGCTGCGTT----- 2206  
Qy 462 AsnLysAlaValLeuTrpPheGlyGlyValProGlyLysGlnTyr-----Ala 478  
Db 2207 -----TTCCGCTTGGCATTCCTCCGCGCGCGAGTATGACGAGCTGGCA 2248  
Qy 479 LeuLeuLysAlaAspGlyIleProLeuGluAsnTyrIleAlaArgThrMetAspGlyVal 498  
Db 2249 AGCTTTTCCAGCAGCGC---ACGGTTGCTCGCTCCTCAAGAGCGGCTGGAAAGCTT 2305  
Qy 499 AlaGluAsnSerHisValArgLysGlnAsnTyrPheTyrTrpAsnCysLeuThrGlyLys 518  
Db 2306 GCTTGCACACTCCCGCTC---ACGGACAATTATTCGCTTGGCAGCGCTTTGCGCGCGCT 2362  
Qy 519 PheLeuArgAspAsn-----CysProThrTyrLeuArgGluAlaAlaPheAlaThr 535  
Db 2363 TATCCGAGCGCATGAGGTGCGCTCGCTTATCTCAAGCGCGAATATTACGAAAG 2422  
Qy 536 LeuLysSerGlyValValAspAsnLeuThrValSerThrAsnPhePheMetGluGluLeu 555  
Db 2423 ATCCGCAACCAACCGCG---CGCGTGGCGGTGCATCAGCCCACTATACCGAGCTGCTT 2479  
Qy 556 Lys-----AlaArgThrTyrThrLysValIleLeuMetAspHisValAspTrpLeu 572  
Db 2480 TCCCGCAAGCGGCAATGGCTGCGACCGGTATATCTGCTGATCGGAGGAGTGGATG 2539  
Qy 573 AspMetProValAlaAsnGluLeuAlaGluCysLeuAlaLysGlnValAlaProGlyGly 592  
Db 2540 ACGGATGTGAGCTCAACAGTATTATGTCGACATCAGCGGCACTGCCGCATCCGGGCA 2599  
Qy 593 IleValIleTrpArgSerAlaSerLeu-----SerProProTyrAlaGlu 607  
Db 2600 CGCGTCATCTTCGCGACCGCGCGCGCA-AAAGAGCGTTATCGAGGGCGCGCTTCGCGCGA 2658  
Qy 608 LeuLeuGlnLysAlaGlyPheAspValArgCysIleArgArgAlaThrGlnGlyTyrMet 627  
Db 2659 CATCCGCAACCGAGTGGGTCTATCTCGAAGAGCGCTCCACGA---ACTCAAC--GCCATG 2713

Qy 628 AspArgValAsnMetTyrSerSerPheTyrMetAlaArgLysGlyAlaLysLys 646  
Db 2714 GACCGCTCGCCATTTATGCGCGCTTCCATATCTACGAGGCGCTATGGCATGAAA 2770

## RESULT 15

BA000012\_13

## WPCOMMENT

Sequence split into 71 fragments LOCUS BA000012 Accession BA000012

Fragment Name	Begin	End
BA000012_00	1	110000
BA000012_01	100001	210000
BA000012_02	200001	310000
BA000012_03	300001	410000
BA000012_04	400001	510000
BA000012_05	500001	610000
BA000012_06	600001	710000
BA000012_07	700001	810000
BA000012_08	800001	910000
BA000012_09	900001	1010000
BA000012_10	1000001	1110000
BA000012_11	1100001	1210000
BA000012_12	1200001	1310000
BA000012_13	1300001	1410000
BA000012_14	1400001	1510000
BA000012_15	1500001	1610000
BA000012_16	1600001	1710000
BA000012_17	1700001	1810000
BA000012_18	1800001	1910000
BA000012_19	1900001	2010000
BA000012_20	2000001	2110000
BA000012_21	2100001	2210000
BA000012_22	2200001	2310000
BA000012_23	2300001	2410000
BA000012_24	2400001	2510000
BA000012_25	2500001	2610000
BA000012_26	2600001	2710000
BA000012_27	2700001	2810000
BA000012_28	2800001	2910000
BA000012_29	2900001	3010000
BA000012_30	3000001	3110000
BA000012_31	3100001	3210000
BA000012_32	3200001	3310000
BA000012_33	3300001	3410000
BA000012_34	3400001	3510000
BA000012_35	3500001	3610000
BA000012_36	3600001	3710000
BA000012_37	3700001	3810000
BA000012_38	3800001	3910000
BA000012_39	3900001	4010000
BA000012_40	4000001	4110000
BA000012_41	4100001	4210000
BA000012_42	4200001	4310000
BA000012_43	4300001	4410000
BA000012_44	4400001	4510000
BA000012_45	4500001	4610000
BA000012_46	4600001	4710000
BA000012_47	4700001	4810000
BA000012_48	4800001	4910000
BA000012_49	4900001	5010000
BA000012_50	5000001	5110000
BA000012_51	5100001	5210000
BA000012_52	5200001	5310000
BA000012_53	5300001	5410000
BA000012_54	5400001	5510000
BA000012_55	5500001	5610000
BA000012_56	5600001	5710000
BA000012_57	5700001	5810000
BA000012_58	5800001	5910000
BA000012_59	5900001	6010000
BA000012_60	6000001	6110000
BA000012_61	6100001	6210000
BA000012_62	6200001	6310000

BA000012\_63 6300001 6410000  
BA000012\_64 6400001 6510000  
BA000012\_65 6500001 6610000  
BA000012\_66 6600001 6710000  
BA000012\_67 6700001 6810000  
BA000012\_68 6800001 6910000  
BA000012\_69 6900001 7010000  
BA000012\_70 7000001 7036071  
Continuation (14 of 71) of BA000012 from base 1300001 (BA000012 Mesorhizobium loti MAF23)

Alignment Scores:  
Pred. No.: 6,31e-17 Length: 110000  
Score: 329.00 Matches: 118  
Percent Similarity: 42.8% Conservative: 66  
Best Local Similarity: 27.4% Mismatches: 184  
Query Match: 9.5% Indels: 63  
DB: 1 Gaps: 15

US-10-620-914-45 (1-648) x BA000012\_13 (1-110000)

Qy 248 TrpIleGlyArgLeuProSer-----ValGlyHisAlaLeuHisGluGluArgValGlu 265  
Db 13624 TGGAAAGCGCGTCTACCAAGCGCGCTTTCCAAAGCGCGCATCTCCGAGCGGTGTT 13683  
Qy 266 ArgProMetPheProPro-ThrPheLeuTyThrGlnSerTrpGluAspProGluPr 285  
Db 13684 CGCCTTCCTGTTTCCGCGCTCGTC-----TATCCGACATCTGGGAAGACCCCGATGT 13737  
Qy 285 oAspMetGluValMetGluLeuAsnProLysAspThrValLeuThrLeuThrSerGlyG1 305  
Db 13738 CGACATGGAGCGCATCGAGCTTGGTCAGGCGCATCGCATCGTCACATCGCTTCGCGCG 13797  
Qy 305 YCysAsnAlaLeuAsnLeuValGlnGlyAlaGlyGlnValValSerValAspCysAs 325  
Db 13798 CTGCACATCTCGCTTACCTACCCGCTTCGCGCGCACGATCGACCGCTCGACCTCAA 13857  
Qy 325 nProAlaGlnSerAlaLeuLeuGluLeuLysValAlaIleGlnGlnLeuGluPheG1 345  
Db 13858 CGCGGCCACATCGCGTGAACCGCATGAAGCTGGAGCGCGTGGCGGCTCTGCCCTCGCA 13917  
Qy 345 u--AspValTrpGlnLeuPheGlyGluGlyValHisProArgIleGluLeuTyTrG1 364  
Db 13918 GGGCGATCTGTCGCTTTTCGGCGCGCGCGCACCGACCAATTCGCAAGCCTATGA 13977  
Qy 364 uLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrpSerLysArgLeuTr 384  
Db 13978 CCGCTTTATTGCGCGCATCTCGATCGCGTTCAGCGCCACTATTGGAGCGCGCAACTG 14037  
Qy 384 p-----TyrPheGlnHisGlyLeuTyTrGlnGlyGlyMetG1 397  
Db 14038 CGGTGTCGGCGGCGATCGCGTCTTCGACCGCAATTTCTACAGACCGCGCTGCTCGG 14097  
Qy 397 YLysLeuCysTrpValLeuGlnCysLeuAlaValValLeuGly----- 411  
Db 14098 C-----CTGTTTCATCGCCATGGCCATCGCAGCGGCAAAATT 14133  
Qy 412 -LeuGlyLysThrValLysArgLeuAlaAsnAlaProThrMetGluGlnArgArgLe 431  
Db 14134 CTTCCGGGTCAACCGCGCCCATGATGGAAGCCGGAATATCGCGCAGCAGCGCGCTT 14193  
Qy 431 uTrpAspSerAsnMetLeuIleHisPheValLysAsnGlyProLysProLeuValTrpLe 451  
Db 14194 CTTCAACAGAGAGTGGCGCGCTTCGACAG-----ANGCTTTGAATGGGC 14244  
Qy 451 uPheValLysPheValSerLeuValLeuPheAsnLysAlaValLeuTrpPheGlyG1 471  
Db 14245 GACCTCGGTAAAGCCTCGCTG-----TTCGCGCTCGG 14277  
Qy 471 YValProGlyLysGlnTyr-----AlaLeuIleLysAlaAspGlyLeProIleGluAs 489  
Db 14278 CATTCGCGCGCGAGTACGATTCCTCATCCTCAGGCGACGCGCATGGCCAGCGT 14337  
Qy 489 nTyIleAlaArg-----ThrMetAspGlyValAlaGluAsnSerHisValAr 505

Db 14338 TCTGAAGCGCGCTGAAAAAGCTCGCCTGCGATTTTCCCTGAAAAAC----- 14386  
Qy 505 gLysGlnAsnTyThrPheTyTrAsnCysLeuThrGlyLysPheLeuArgAspAsn----- 523  
Db 14387 -----AATTATTTCGCTCGCAGGCTTTTGGCCCGCGCTATCCAAATCCCGGTGAGGC 14439  
Qy 524 ----CysProThrTyLeuArgGluAlaAlaPheAlaThrLeuLysSerGlyValValAs 542  
Db 14440 CGCCCTGCGCGCTATCTGAAAGAGAGAACTACGAAACCATCCGC---GGCAATATCGA 14496  
Qy 542 pAsnLeuThrValSerThrAsnPhePheMetGluGluLeu-----LysAlaArgTh 559  
Db 14497 CGCGTTCGCGCATCCACCATGCCAATCTGATCGAATTCCTCGCGCAAGACGCGGCAC 14556  
Qy 559 rTyThrLysValIleLeuMetAspHisValAspTrpLeuAspMetProValAlaAsnG1 579  
Db 14557 CGTCGATCGCTTCATCTGCTCGATCGCAGGAGTGGATGACCGATGACCGCTCAACGC 14616  
Qy 579 uLeuAlaGluCysLeuAlaLysGlnValAlaProGlyGlyIleValIleTrpArgSerAl 599  
Db 14617 GCTGTGTCGGAATCAGCCGACCGCTCCGCGAGCGCGCGCTCATCTTCGCGACCGC 14676  
Qy 599 a-----SerLeuSerProProTyTrAlaGluLeuIleGlnLysAlaGlyPheAsp-V 616  
Db 14677 CGCGGAGCGCGAGCTGTGTCGCGCGCGCTCTCGACCTCGCTCGACCGTGGGACTA 14736  
Qy 616 alArgCysIleArgArgAlaThrGlnGlyTyTrMetAspArgValAsnMetTyTrSerSerP 636  
Db 14737 TCAGGACGAGCGCTCGCGGAAATTCGCGCACGC-GACCGTTCGGCCATCTATGGCGGCT 14795  
Qy 636 heTyrMetAlaArgArgLysGlyAla 644  
Db 14796 TCACCTCTATGTGAAGCGCACCGCA 14821

Search completed: March 14, 2006, 05:00:33  
Job time : 9713 secs



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 13, 2006, 21:06:21 ; Search time 9929 Seconds  
(without alignments)  
11146.561 Million cell updates/sec

Title: US-10-620-914-44

Perfect score: 1947

Sequence: 1 atggggctcggtcgctgacgg.....ggccgaagaaggacaactaa 1947

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_in:\*  
3: gb\_env:\*  
4: gb\_sm:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pr:\*  
9: gb\_ro:\*  
10: gb\_sts:\*  
11: gb\_sy:\*  
12: gb\_un:\*  
13: gb\_vi:\*  
14: gb\_htg:\*  
15: gb\_pl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1883	96.7	2001	15	AY656806 Chlamydom
2	100.6	5.2	110000	15	Continuation (13 o
3	100.2	5.1	3045	1	AF329857 Rhodobact
4	87.4	4.5	110000	1	Continuation (17 o
5	83.8	4.3	110000	15	Continuation (21 o
6	82.2	4.2	125020	8	AF429315 Homo sapi
7	82	4.2	31065	15	AC152119 Pichia st
8	80	4.1	110000	1	Continuation (64 o
9	77.2	4.0	110000	15	Continuation (21 o
10	76.8	3.9	349260	1	BX572595 Rhodosphe
11	76	3.9	1932	1	AJ605554 Nonomurae
12	75.6	3.9	137241	13	AY386263 Orf virus
13	74	3.8	110000	1	Continuation (20 o
14	73	3.7	75337	15	BX842626 Neurospor
15	72.6	3.7	110000	1	Continuation (25 o
16	72.4	3.7	137560	6	CQ830057 Sequence
17	72.4	3.7	137560	6	AX754989 Sequence
18	70.8	3.6	110000	1	Continuation (14 o

c	19	70.4	3.6	314100	1	SC0939106	AL939106 Streptomy
	20	70	3.6	294800	1	SMES91789	AL591789 Sinorhizo
	21	69.8	3.6	305584	1	AE016920	AE016920 Chromobac
c	22	69.6	3.6	300100	1	SC0939123	AL939123 Streptomy
	23	69.4	3.6	12370	1	AE004680	AE004680 Pseudomon
c	24	69.4	3.6	189050	1	AL646086	AL646086 Ralstonia
	25	69.2	3.6	1500	13	AF294740	AF294740 Baboon he
	26	69	3.5	2563	1	AY228174S1	AY228174 Streptomy
	27	68.6	3.5	110000	14	CT005248_1	Continuation (2 of
	28	68.4	3.5	66669	1	AME16952	Y16952 Amycolatops
	29	68.2	3.5	2440	1	SLGROESEL	X95970 S. lividans
	30	68.2	3.5	2830	2	AF027972	AF027972 Nephila c
	31	68.2	3.5	2830	6	AR091362	AR091362 Sequence
	32	67.8	3.5	1092	1	AB046667	AB046667 Streptomy
	33	67.8	3.5	1092	6	BD095293	BD095293 Genes of
	34	67.8	3.5	1092	6	BD141256	BD141256 Isopenten
	35	67.8	3.5	1092	6	E48856	E48856 Gene of enz
	36	67.8	3.5	1104	6	BD097684	BD097684 A method
	37	67.8	3.5	1104	6	BD141277	BD141277 Isopenten
	38	67.8	3.5	6798	1	AB037666	AB037666 Streptomy
	39	67.8	3.5	6798	6	BD095289	BD095289 Genes of
	40	67.8	3.5	6798	6	BD097685	BD097685 A method
	41	67.8	3.5	6798	6	E48852	E48852 Gene of enz
	42	67.8	3.5	30837	1	AJ843080	AJ843080 Streptomy
	43	67.8	3.5	39237	1	AB211959	AB211959 Streptomy
c	44	67.8	3.5	42149	1	AJ786317	AJ786317 Streptomy
	45	67.8	3.5	139962	13	AY386264	AY386264 Orf virus

## ALIGNMENTS

RESULT 1	AY656806	2001 bp	mrna	linear	PLN 11-FEB-2005
LOCUS	Chlamydomonas reinhardtii betaine lipid synthase (BTA1) mRNA,				
DEFINITION	complete cds.				
ACCESSION	AY656806				
VERSION	AY656806.1	GI:50261571			
KEYWORDS	Chlamydomonas reinhardtii				
SOURCE	Chlamydomonas reinhardtii				
ORGANISM	Chlamydomonas reinhardtii				
REFERENCE	1 (bases 1 to 2001)				
AUTHORS	Riekhof, W.R., Sears, B.B. and Benning, C.				
TITLE	Annotation of Genes Involved in Glycerolipid Biosynthesis in Chlamydomonas reinhardtii: Discovery of the Betaine Lipid Synthase				
JOURNAL	BTA1Cr				
PUBMED	Eukaryotic Cell 4 (2), 242-252 (2005)				
REFERENCE	2 (bases 1 to 2001)				
AUTHORS	Riekhof, W.R. and Benning, C.				
TITLE	Direct Submission				
JOURNAL	Submitted (17-JUN-2004) BMB, MSU, Biochemistry 215, East Lansing, MI 48824-1319, USA				
FEATURES	Location/Qualifiers				
source	1..2001				
	/organism="Chlamydomonas reinhardtii"				
	/mol_type="mRNA"				
	/strain="CC125"				
gene	1..2001				
	/gene="BTA1"				
CDS	1..2001				
	/gene="BTA1"				
	/note="multifunctional protein with S-adenosylmethionine:diacylglycerol 3-amino-3-carboxypropyl transferase and S-adenosylmethionine:diacylglycerolhomoserine N-methyltransferase activities"				
	/codon_start=1				
	/product="betaine lipid synthase"				

/protein_id="AA72342.1"		Db		781		GCCCCCTACTACTGTTGGATTGGCCGCTGCCAGCGTTGGCCACGCGCTTCACGAGGAG		840	
/db_xref="GI:50261572"		Qy		787		CGCGTGGAGGGCGGCCCACTGTTCCCGCCCACTCTCTCTACACGCACTGCTGGAGGAC		846	
/translation="MSGRDGRPASVTKNFKLSKLSKMDLLTLVRHWFSGKKG		Db		841		CGCGTGGAGGGCGGCCCACTGTTCCCGCCCACTCTCTCTACACGCACTGCTGGAGGAC		900	
DHAARLESFYGQAAYDAFRSRFLNWRPMLAAVARLAERSNLIWDLGGGTGEN		Qy		847		CCCGAGCCGATATGAGAGTGATGAGATCAACCCCAAGAGACAGGCTGCTGACCCCTGACT		906	
VDMADYIDIAKFKSYVDLCHSLCEVAKKAKAGKRNQVVEADACQFAPPEGTA		Db		901		CCCGAGCCGATATGAGAGTGATGAGATCAACCCCAAGAGACAGGCTGCTGACCCCTGACT		960	
TLITFSYSLTMIPIFNHVIDQAGSYLSQDLGVADFYSGKYDLPLRQMPWRFFFW		Qy		907		AGCGGCGGTGCAATCCCTGAACTCTGCTGAGAGGGGCGCGCAGAGTGTCTGGTG		966	
RSIFDINDIDIGPERAYLEQKRWQNTQGISIPVPMWRAPYVWIGRLPSVGH		Db		961		AGCGGCGGTGCAATCCCTGAACTCTGCTGAGAGGGGCGCGCAGAGTGTCTGGTG		1020	
LHEVERPMPFPPTFLYTQSWEDPDPMEVINEPKDVTLTITSGCNALNLLVOGA		Qy		967		GACTGCAACCCCGCGCAGTCCGCGCTCTGAGCTGAAGAGTGGCCATTCTACGAGCTG		1026	
GOVVSVDCAQASALLEKVAIQOLEPFDVWOLFEGVHPRLEELYEKLAPELSQT		Db		1021		GACTGCAACCCCGCGCAGTCCGCGCTCTGAGCTGAAGAGTGGCCATTCTACGAGCTG		1080	
SHNFKRLMTFQHLGYQGMGKLVLCVLAQLVGLGKTVKRLANAIPMESQRILW		Qy		1027		GAGTTTGAAGACGCTGGCAGCTGTTTCGGCGAGGGCGTGCAACCCGCGCATTTGAGGAGCTG		1086	
DSNMLHFVNGKPLVLFVKSFLVLFNKAIVLFGGVPFGVQYALIKADGIPIN		Db		1081		GAGTTTGAAGACGCTGGCAGCTGTTTCGGCGAGGGCGTGCAACCCGCGCATTTGAGGAGCTG		1140	
IARTMDGVAENSHVRQNYFYINCLTGKFLRDNCPYLRREAFATLKSGVVDNLTYST		Qy		1087		TACGAGAAGAAGCTGGCGCCCTTCTCTGTCGCAAAACAGCCACAATCTCTGGTCCAAAGCGC		1146	
NFWBELKATYTKVILMDHVDLMDPMEVANECLAQVAPGGIVIMRSASLSPPYA		Db		1141		TACGAGAAGAAGCTGGCGCCCTTCTCTGTCGCAAAACAGCCACAATCTCTGGTCCAAAGCGC		1200	
ELIQKAGFVDCIRRTAQGYMDRVNMYSSPYMARRKGAKDN"		Qy		1147		CTCTGTGATTTCCAGACAGCGCTGTACTACAGGGGGGCGCATTTGAGGAGCTG		1206	
ORIGIN		Db		1201		CTCTGTGATTTCCAGACAGCGCTGTACTACAGGGGGGCGCATTTGAGGAGCTG		1260	
Query Match 96.7%; Score 1883; DB 15; Length 2001;		Qy		1207		CTCAGTGTCTGGCGCTGTGGTGTGGGCAAGACCGCTCAAGGGCTCTGCGCAACGCG		1266	
Best Local Similarity 97.3%; Pred. No. 1.4e-210;		Db		1261		CTCAGTGTCTGGCGCTGTGGTGTGGGCAAGACCGCTCAAGGGCTCTGCGCAACGCG		1320	
Matches 1947; Conservative 0; Mismatches 0; Indels 54; Gaps 1;		Qy		1267		CCCAAAATGAGAGAGCAGCGCCCTGTTGTGGGAACAGCAATGCTCATCTTCTGTGAAG		1326	
1 ATGGGGTCGGGTGCTGAGCGCGCGCTGCGAGCTACACCAAGAGAACTTCTCCCTGGAG		Db		1321		CCCAAAATGAGAGAGCAGCGCCCTGTTGTGGGAACAGCAATGCTCATCTTCTGTGAAG		1380	
61 AAGCTCAAGCTCAGCAGATGAAGATGACTGACCGTTCTGGCCATATATGTGTTCCGC		Qy		1327		AACGGGCCCAAGCGCTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGT		1386	
61 AAGCTCAAGCTCAGCAGATGAAGATGACTGACCGTTCTGGCCATATATGTGTTCCGC		Db		1381		AACGGGCCCAAGCGCTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGT		1440	
121 AGCAAGAAGCGGATGATACGCTGCTCGCTGAGAGCTTCTACGGGCGCCAGGCGCT		Qy		1387		AAGCGCGT		1446	
121 AGCAAGAAGCGGATGATACGCTGCTCGCTGAGAGCTTCTACGGGCGCCAGGCGCT		Db		1441		AAGCGCGT		1500	
181 GCCT-----TT		Qy		1447		GACGGCATCCCAATGAGAACTACATCGCGCGCAACATGGAACGGCGTGGCGGAACTCG		1506	
181 GCCTATGATGTTTCGGTCCCGTTCCTCTGGGGTCCGAGGCCATGCTGCTGTCAGTT		Db		1501		GACGGCATCCCAATGAGAACTACATCGCGCGCAACATGGAACGGCGTGGCGGAACTCG		1560	
187 GCTCCCGCTGGCGAGCGCTCGAACTCATCTGGGTGACCTGGGTGGTGGCACTGGG		Qy		1507		CAGTGGCGCAAGCAGAACTACTTCTACTACAACTGCTCAACCGCAAGTTCTCTGCGGAC		1566	
241 GCTGCCCGCTGGCGAGCGCTCGAACTCATCTGGGTGACCTGGGTGGTGGCACTGGG		Db		1561		CAGTGGCGCAAGCAGAACTACTTCTACTACAACTGCTCAACCGCAAGTTCTCTGCGGAC		1620	
247 GAGAATGCGATATGATGGCTGATTAATCATCGACTGCGGAAGTTCAAGTCCATCTACGTG		Qy		1567		AACTGCCCACTTACCTGCGGAGGGCGCTTCCCAACCTCAAGAGTGGCGTGGTGGAC		1626	
301 GAGAATGCGATATGATGGCTGATTAATCATCGACTGCGGAAGTTCAAGTCCATCTACGTG		Db		1621		AACTGCCCACTTACCTGCGGAGGGCGCTTCCCAACCTCAAGAGTGGCGTGGTGGAC		1680	
307 GTCCAGCTGCGCTGCTGCGAGTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG		Qy		1627		AACTTGACCGTCTCCACCAACTTCTTATGAGAGAGCTCAAAAGCGCGCACTACCAAG		1686	
361 GTCCAGCTGCGCTGCTGCGAGTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG		Db		1681		AACTTGACCGTCTCCACCAACTTCTTATGAGAGAGCTCAAAAGCGCGCACTACCAAG		1740	
367 AAGAAATGTCAGGTGCTGAGGCGCGCTGCGCAATTTGCGCCCTGAGGGCACCGCG		Qy		1687		GTGATTTGATGAGACCACTGAGTGGCTGGATATGCCCTGGCGCAACGAGCTGGCGGAG		1746	
421 AAGAAATGTCAGGTGCTGAGGCGCGCTGCGCAATTTGCGCCCTGAGGGCACCGCG		Db		1741		GTGATTTGATGAGACCACTGAGTGGCTGGATATGCCCTGGCGCAACGAGCTGGCGGAG		1800	
427 ACCTCATACCTTCTCTACTGCTCAAGATGATTCACCGTTTCCAACAGTCAATCGAC		Qy		1747		TGCCTGGCAAGCAGGTTGCGCGGGCGGCTGCTCATCTTGGGCGCTCCGCTCCCTCAGC		1806	
481 ACCTCATACCTTCTCTACTGCTCAAGATGATTCACCGTTTCCAACAGTCAATCGAC		Db		1801		TGCCTGGCAAGCAGGTTGCGCGGGCGGCTGCTCATCTTGGGCGCTCCGCTCCCTCAGC		1860	
487 CAGCTTGTCTGCTTCTGCTCCAGAGCGGCTGCTGGCGTGGCGGCTTCTACGTGAGC		Qy		1807		CCGCGCTTACGCGGAGCTGATTCAGAGAGCGGGCTTCGACGTGCGCTGCTGCTCGCGGCC		1866	
541 CAGGCTTGTCTGCTTCTGCTCCAGAGCGGCTGCTGGCGGCTTCTACGTGAGC		Db							
547 GGCAAGTACGACCTGCGCCCTGCGCAGATGCTGCTGCGCGCTTCTCTGCGGATCG		Qy							
601 GGCAAGTACGACCTGCGCCCTGCGCAGATGCTGCTGCGCGCTTCTCTGCGGATCG		Db							
607 ATCTTTCGACATCGAACAATTGACATCGGCGCCGAGCGCGCGCTTACCTGGAGCAAG		Qy							
661 ATCTTTCGACATCGAACAATTGACATCGGCGCCGAGCGCGCGCTTACCTGGAGCAAG		Db							
667 CTGAGCGCGTGTGGGAGCAGAACACCCAGGGTTCGATCCCTACGTCGGTGGCTGGC		Qy							
721 CTGAGCGCGTGTGGGAGCAGAACACCCAGGGTTCGATCCCTACGTCGGTGGCTGGC		Db							
727 GCGCCCTACTACGTGTGGATTGGCGCGCTGCGCCAGCGTTGGCCACGCGCTTGCACGAGGAG		Qy							

Db 1861 CCGCCCTACCCGAGCTGATCCAGAGCGGGCTTCGACGTGGCTGCATCCCGCGGCC 1920  
 QY 1867 ACTCAGGGCTACATGGACCGCTCAACATGTACAGCTCTCTTACATGGCCCGCCGGAAG 1926  
 Db 1921 ACTCAGGGCTACATGGACCGCTCAACATGTACAGCTCTCTTACATGGCCCGCCGGAAG 1980  
 QY 1927 GCGCCCAAGAGGACAACTAA 1947  
 Db 1981 GCGCCCAAGAGGACAACTAA 2001

## RESULT 2

CR382131\_12

WPCOMMENT

Sequence split into 43 fragments LOCUS CR382131 Accession CR382131

Fragment Name	Begin	End
CR382131_00	1	110000
CR382131_01	100001	210000
CR382131_02	200001	310000
CR382131_03	300001	410000
CR382131_04	400001	510000
CR382131_05	500001	610000
CR382131_06	600001	710000
CR382131_07	700001	810000
CR382131_08	800001	910000
CR382131_09	900001	1010000
CR382131_10	1000001	1110000
CR382131_11	1100001	1210000
CR382131_12	1200001	1310000
CR382131_13	1300001	1410000
CR382131_14	1400001	1510000
CR382131_15	1500001	1610000
CR382131_16	1600001	1710000
CR382131_17	1700001	1810000
CR382131_18	1800001	1910000
CR382131_19	1900001	2010000
CR382131_20	2000001	2110000
CR382131_21	2100001	2210000
CR382131_22	2200001	2310000
CR382131_23	2300001	2410000
CR382131_24	2400001	2510000
CR382131_25	2500001	2610000
CR382131_26	2600001	2710000
CR382131_27	2700001	2810000
CR382131_28	2800001	2910000
CR382131_29	2900001	3010000
CR382131_30	3000001	3110000
CR382131_31	3100001	3210000
CR382131_32	3200001	3310000
CR382131_33	3300001	3410000
CR382131_34	3400001	3510000
CR382131_35	3500001	3610000
CR382131_36	3600001	3710000
CR382131_37	3700001	3810000
CR382131_38	3800001	3910000
CR382131_39	3900001	4010000
CR382131_40	4000001	4110000
CR382131_41	4100001	4210000
CR382131_42	4200001	4224103

Continuation (13 of 43) of CR382131 from base 1200001 (CR382131 Yarrowia lipolytica chr

Query Match 5.2%; Score 100.6; DB 15; Length 110000;  
 Best Local Similarity 58.2%; Pred. No. 0.0034;  
 Matches 216; Conservative 0; Mismatches 149; Indels 6; Gaps 2;

QY 178 GCTGCTTTGTCGCCCGCTGGCGGAGCGCTCGAACCTCATCTGGGTTGACCTGGGTGT 237  
 Db 59775 GCACACCGCGCGCGAGCTTCGAAGAAGCATGATCTCGTGGGTCGACATTTGAGGA 59834

QY 238 GGCACCTG---GGGAGAAATGCGATATGATGGCTGATTACATGACCTGGCGAAGTTCAAG 294  
 Db 59835 GGAACCGGCTGGAACTTGAACATGAGCACATGACCACTCTTCCCATTTCTCGCTTCAAG 59894

QY 295 TCATCTACGTGCTGACCTGTGCCACTCGCTGTGCGAGGTGGCCCAAGAAGAGGCGAAG 354  
 Db 59895 GCCATCTACCTGTTGATCTGAGTCTTCTCTGTGCGAGGTGGCCCGAAGCGATTTCGCC 59954  
 QY 355 GCCAAGGGCTGGAAGATGTCAGGTCGTGGAGCGCGCTGCGCAATTGCGCCCTT 414  
 Db 59955 GAGAAGACTGGAAGAACGTGCACGTTCTGTGCATCGAGCTGCCGATTTCATCTCCCTCC 60013  
 QY 415 GAGGG---CACCGCGACGCTCATCCTTCTCTCTACTCGCTCACGATGATTCACCGCTTC 471  
 Db 60015 GCTGGATCTCTCGTTGATCTCTTCCACCATGCTCTCTCTCTATGATCCCACTAC 60074  
 QY 472 CACACGCTCATCGACCGCTTCTGCTGCTACCTGTGTCACAGAGCGCTGGTGGCGGTGCC 531  
 Db 60075 TATGCCGTTATCGATCGACTCTCCACCTGCTTGGCAGGACGCTCTGCTCACTGTCATC 60134  
 QY 532 GACTTCTACGT 542  
 Db 60135 GACTTTTACGT 60145

## RESULT 3

AF329857

LOCUS

DEFINITION

AF329857

ACCESSION

AF329857

VERSION

AF329857.1

KEYWORDS

SOURCE

ORGANISM

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

3045 bp DNA linear BCT 16-MAY-2001  
 Rhodospirillum rubrum betaine lipid biosynthesis operon, complete sequence.

AF329857

AF329857.1

GI:14091029

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

Rhodospirillum rubrum

1 (bases 1 to 3045)  
 Klug, R.M. and Benning, C.  
 Two enzymes of diacylglycerol-O'-4'-(N,N,N-trimethyl)homoserine biosynthesis are encoded by btaA and btaB in the purple bacterium Rhodospirillum rubrum

Proc. Natl. Acad. Sci. U.S.A. 98 (10), 5910-5915 (2001)

11311765

2 (bases 1 to 3045)

Klug, R.M. and Benning, C.

Direct Submission

Submitted (18-DEC-2000) Department of Biochemistry &amp; Molecular

Biology, Michigan State University, 224 Biochemistry Bldg., East

Lansing, MI 48824-1319, USA

Location/Qualifiers

1..3045

/organism="Rhodospirillum rubrum"

/mol\_type="genomic DNA"

/strain="2.4.1"

/db\_xref="taxon:1063"

544..2423

/gene="betaine lipid biosynthesis operon"

544..1794

/gene="btaA"

544..1794

/gene="btaB"

/codon\_start=1

/transl\_table=11

/product="S-adenosylmethionine:diacylglycerol

3-amino-3-carboxypropyl transferase"

/protein\_id="AAK53560.1"

/db\_xref="GI:14091030"

/translation="MTQFALTHLPAPVARIQAAVHPTSLASGLMERMSRLPHG

LVPTQWSDPAVDMAALIRPGDLRLVAISGNCNVLSTLQSGFSLAVDLSPAHVAL

GRLLKAARTLPDHAFFDLFRADLPNGAALYDRHIALPDGRSRRYWEASPFRR

IQLFERGYRICALGRFTCAATLARAAGTDLRGFLDCPDIAQRSPFYAHIGLPFA

PVQALARRPALFGLIPPAQYALLAGDGDVLPVLRHLRLICDPLRENYPAF

QAARYPYRPGEGALPPYLEPTAFETLRENAGRVQIENRSLTEALAAPEESIHGPTL

LDAQWTDALQLTALRWQVTRTAAPGARVIFPTGGADLLPGRVPEEIIHWRADRA

GAGHAARSALYGGFHLRRDA

1791..2423

gene



Qy	980	CGCAGTCGGCGCTTCCTGGAGCTCAAGAACGGTGGCCATTACAGACGTGGAGTTTGAGGACG	1039
Db	27488	GGCAGACCTACCTTCGTGGAACTCAAGCTGGCGCCATCCGGCACTTCTCTACGACACCC	27547
Qy	1040	TGTGGCAGCTGTTTCGGCGAGGGCGTCACCCGCGATTGAGGAGCTGTACGAGAGAAGC	1099
Db	27548	TGCTCGAGTTCCTCGGGTGCACCCGAGCCGGGACCGCTGGCGGCTGTTCG---AGGAGC	27604
Qy	1100	TGGCGCCCTTCCTGTTCGCAAAACGAGCACAACCTTCGTGTCCAAGCGCCTCTGGTACTTCC	1159
Db	27605	TCACCCCGCGCTGTCAACCGGCGCGTCGCCTATTTCAACCGCTACCACAAGGCGATTCC	27664
Qy	1160	AGCACCGCTGTACTACTACAGGGGGGATGGCAAGCTGTCTGGGTGTCTGAGTGCCTGG	1219
Db	27665	GGGTGGCGCTGTCTGTCTGGCGCGGCGACACGAGCGGCTCTACGTGGCGGCTGTGGCGCGCA	27724
Qy	1220	CGCTGTGTGGGACTGGGCAAGACGGTCAAGCGCTGCCAAACGCGCCCAATATGGAGG	1279
Db	27725	TGATGGCAGCTCTACGGCTCGGCGATGAGGAACCTCTTCGCCCGCGCGACCTCGAGG	27784
Qy	1280	AGCAGCGCGCTCTGTGGGACAGCAACATCTCTCATCCACTTCGTGAAGAACGGGCGCCCAAGC	1339
Db	27785	CGCAGCGCGCGTGT-----ACGGGAGAAGATCGACGGGTGTCTGTGGCGCAGC	27835
Qy	1340	CGCTGTGTGGCTGTTCGTCAAGTTCGTGAGCCTGGTGTCTTTCAAAGGCGGTGTGT	1399
Db	27836	TGATCGCGGCGGGTTTCAACCGAACGACCCCTCAAGACGGTGTCAACGACGACTCTACA	27895
Qy	1400	GGTTCGGCGCGCGTGCGGGGCAAGCAGTACCGCTGATCAAGCGGAGCGGCATCCCCA	1459
Db	27896	ACGTCTGTACCGATGTGCAGTCTGTGGGGGATCTACGTGTCTGAACGGTCTGAGACACCC	27955
Qy	1460	TTTGAGAACTACATCGCGCGCACCATGAGCGGCGTGGCGAGAACTTCGCACCTGGCGAAGC	1519
Db	27956	TCACCGAGCACCTGTGTCGCGGCAACGACTGGGTGTCTCATGTCTGACGGCGGCTACC	28015
Qy	1520	AGAAC	1524
Db	28016	CGGAC	28020

## RESULT 5

CR382138 20/c

WPCOMMENT\_

Sequence split into 24 fragments LOCUS CR382138 Accession CR382138

Fragment Name	Begin	End
CR382138_00	1	110000
CR382138_01	100001	210000
CR382138_02	200001	310000
CR382138_03	300001	410000
CR382138_04	400001	510000
CR382138_05	500001	610000
CR382138_06	600001	710000
CR382138_07	700001	810000
CR382138_08	800001	910000
CR382138_09	900001	1010000
CR382138_10	1000001	1110000
CR382138_11	1100001	1210000
CR382138_12	1200001	1310000
CR382138_13	1300001	1410000
CR382138_14	1400001	1510000
CR382138_15	1500001	1610000
CR382138_16	1600001	1710000
CR382138_17	1700001	1810000
CR382138_18	1800001	1910000
CR382138_19	1900001	2010000
CR382138_20	2000001	2110000
CR382138_21	2100001	2210000
CR382138_22	2200001	2310000
CR382138_23	2300001	2336804

### Query Match

4.38; Score 83.8; DB 15; Length 110000;

Best Local Similarity 54.0%; Pred. No. 0.31; Matches 194; Conservative				Mismatches 162; Indels 3; Gaps 1;			
Qy	188	CTGCCCGCTGGCCGAGCGCTCGAACCTCATCTGGGTTGACCTGGGTGGTGGCACTGGGG	247				
Db	59864	CCGACATCTCTTAAGAAGAAGATCTTGTTGGGTAGACATTTGGAGGGGGAAGTGGCT	58805				
Qy	248	AGAATCTGCATATGATGGCTGATTAATCATCGACCTGGC---GAAGTTCAAGTCCATCTACG	304				
Db	59804	CCAATATCGATGATACATGACACGAGTAATGTGCGATCTCCAAGAAATTTCAAGCTGTTATC	58745				
Qy	305	TGGTCGACCTTGCCACTCGCTGTCGGAGGTGGCCAGAAGAGCGAGGCCCAAGGCT	364				
Db	58744	TTGTGATTTATACCATCTGTTGTCGGAAGTTGCGAAAAAAGGTTCCGCTGATGAAGAGAT	58685				
Qy	365	GGAAAGAAATCTCCAGGTGCTGGAGGCGGACGCTTGGCAATTTGCGCCCCCTCAGGCGCACCG	424				
Db	59684	GGAGCAATGTGATGTTCTAGTTGGAGATGCTTGTGATTTCCGTTATGGGTATGAAGAAG	58625				
Qy	425	CGAGCGCTCATCACTTCCTCTCTACTCGCTCAAGATGATTCACCGTTCCAAACGTCATCG	484				
Db	58624	CCGATTTGGTGACCTTCCTACTCATTTGTCTATGATTTCCAACTTTTCCATGCTGCTGTG	58565				
Qy	485	ACCAGGCTTGTCTGATACCTGTCCTCCAGAGCGGCTGGTGGGTTGCCGCACTTCTACGTG	543				
Db	58564	ATCATGCTCTCAAGTTATTAGACAAACAGGTTGTTATCGCCCTGTGTAGACTTCGGTGTG	58506				
RESULT 6							
AF429315/c							
LOCUS							
DEFINITION Homo sapiens junctophilin 3 (JPH3) gene, partial cds.							
ACCESSION AF429315							
VERSION AF429315.1 GI:17646244							
KEYWORDS							
SOURCE Homo sapiens (human)							
ORGANISM Homo sapiens							
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;							
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;							
Hominoidea; Homo.							
REFERENCE							
AUTHORS							
Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,							
Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,							
Potter,N.T., Rose,C.A. and Margolis,R.L.							
A repeat expansion in the gene encoding junctophilin-3 is							
associated with Huntington disease-like 2							
Nat. Genet. 29 (4), 377-378 (2001)							
JOURNAL							
PUBMED							
11694876							
REFERENCE							
2 (bases 1 to 125020)							
Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.							
Direct Submission							
TITLE							
Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical							
Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA							
FEATURES							
source							
1..125020							
/organism="Homo sapiens"							
/mol_type="genomic DNA"							
/db_xref="taxon:9606"							
/chromosome="16"							
/map="16q24.3; between D16S20 and WI-12410"							
/note="Isolated from a patient with Huntington's							
Disease-Like 2 (HDL2)"							
complement(35581..35746)							
repeat_region							
/rpt_type=tandem							
/rpt_unit="ctg"							
complement(<36507..>36887)							
gene							
/gene="JPH3"							
/note="synonym: JP3"							
complement(<36507..>36887)							
mRNA							
/gene="JPH3"							
/product="junctophilin 3"							
complement(<36507..>36887)							
CDS							
/gene="JPH3"							





QY 395 CTTCGCAATTTGGCCCTGAGGGGACCGGAGGCTCATCACCTTCTCTACTGCTCA 454  
DB 20341 CCTGTGACTTTGAATCGACTACAGAGCTGCTGACTTGATCAGCTTTTCTACTCTCTT 20400  
QY 455 CGATGATTCCACGTTCCCAACGTCATCGACGAGGCTTCTGCTACTCTCCAGAGC 514  
DB 20401 CTATGATTCCACTTTCAATGCTGCTATCGAATGATGATCACCAGTTGATAGCAGG 20460  
QY 515 GCCTGTGGGCGTTGCCGACTT 536  
DB 20461 GTATTATTGCTGTAGACTT 20482

RESULT 8  
BA000040\_63/c  
WPCOMMENT

Sequence split into 91 fragments LOCUS BA000040 Accession BA000040

Fragment Name	Begin	End
BA000040_00	1	110000
BA000040_01	100001	210000
BA000040_02	200001	310000
BA000040_03	300001	410000
BA000040_04	400001	510000
BA000040_05	500001	610000
BA000040_06	600001	710000
BA000040_07	700001	810000
BA000040_08	800001	910000
BA000040_09	900001	1010000
BA000040_10	1000001	1110000
BA000040_11	1100001	1210000
BA000040_12	1200001	1310000
BA000040_13	1300001	1410000
BA000040_14	1400001	1510000
BA000040_15	1500001	1610000
BA000040_16	1600001	1710000
BA000040_17	1700001	1810000
BA000040_18	1800001	1910000
BA000040_19	1900001	2010000
BA000040_20	2000001	2110000
BA000040_21	2100001	2210000
BA000040_22	2200001	2310000
BA000040_23	2300001	2410000
BA000040_24	2400001	2510000
BA000040_25	2500001	2610000
BA000040_26	2600001	2710000
BA000040_27	2700001	2810000
BA000040_28	2800001	2910000
BA000040_29	2900001	3010000
BA000040_30	3000001	3110000
BA000040_31	3100001	3210000
BA000040_32	3200001	3310000
BA000040_33	3300001	3410000
BA000040_34	3400001	3510000
BA000040_35	3500001	3610000
BA000040_36	3600001	3710000
BA000040_37	3700001	3810000
BA000040_38	3800001	3910000
BA000040_39	3900001	4010000
BA000040_40	4000001	4110000
BA000040_41	4100001	4210000
BA000040_42	4200001	4310000
BA000040_43	4300001	4410000
BA000040_44	4400001	4510000
BA000040_45	4500001	4610000
BA000040_46	4600001	4710000
BA000040_47	4700001	4810000
BA000040_48	4800001	4910000
BA000040_49	4900001	5010000
BA000040_50	5000001	5110000
BA000040_51	5100001	5210000
BA000040_52	5200001	5310000
BA000040_53	5300001	5410000
BA000040_54	5400001	5510000

Continuation (64 of 91) of BA000040 from base 6300001 (BA000040 Bradyrhizobium japonicum

Query Match 4.1%; Score 80; DB 1; Length 110000;  
Best Local Similarity 45.6%; Pred. No. 0.86;  
Matches 379; Conservative 0; Mismatches 435; Indels 18; Gaps 2;

QY	972	CAACCCGCGAGTCGCGCTTCTGAGCTGAAGAAGTGGCCATTCAGCAGCTGGAGTT	1031
DB	73879	CATGCTCGAGACCACCGCGCGCTGGAGAACCAAGATCAACAGTCGCGAGAAGAT	73820
QY	1032	TGAGGACGTTGCGAGCTGTTGCGGAGGGGCTGACCCGCGCATTTGAGGAGCTGTACGA	1091
DB	73819	GGAGACGGTCGCGCAGCTCGCGCGCGCATTCGCCACAGCTTCAACAACGTCGTGTCGC	73760
QY	1092	GAAGAAGCTGGCGCCCTTCTGTGCAAAACCAAGCCACAACTTCTGTCCAAAGCGCCTGTG	1151
DB	73759	CATCATGATGGCGAAGCATTTCTGTGAACGGCGCACAGCCGACCGATCCGTCGTCCA	73700
QY	1152	GTACTTCCAGCAGCGCTGTACTACAGGGGCGCATGGGCAAGCTGTGTGGTGTGCA	1211
DB	73699	CGACATCATGCAGTCAAGCAGAACCGACGCGCGCGACGCTGTGTGGCGAGCTGCT	73640
QY	1212	GTGCTGGCGGTGTGCTGGAGTGGCAAGACCGTCAAGCGCTCGCCAAAGCGC---C	1268
DB	73639	GGCGTTCTCGCGCGCGCAGACGCTGCGCCGCGAGGTGCTCGATCTCGCGATCGCTTTC	73580
QY	1269	CACAATGGAGGAGCAGCGCGCTGTGTGGGACAGCAACATGCTCATCCACTTCGTGAAGAA	1328
DB	73579	CGACCTCACCATGCTGCTGCGCGGCTGATCGCGAGAGGTCAAGCTGACCTGATCCA	73520
QY	1329	CGGGCCCAAGCGCTGGTGTGCTGTTGTCGAAGTTGTTGAGCCTGGTGTCTTTCACAA	1388
DB	73519	CGGCGCGACCTCTGCGCGGTGAAGTGTGACGCTCTCCAGTTCGAGAGGTGATCGTCAA	73460
QY	1389	GGCGGTGCTGTGTTGCGCGCGGCTGCGGGCAAGCAGTACGCTGTATCAAGCGGA	1448
DB	73459	TCTCGCGGTGAACGCGCCCGACGCGCATGCCCGACGCGCGCAAGCTGATCATCGCACCG	73400



```

misc_feature
/locus tag="RPA0632"
/notes="4 probable transmembrane helices predicted
byTMHMM2.0"
complement (2036..2377)
/genes="rnpA"
/locus tag="RPA0633"
complement (2036..2377)
/genes="rnpA"
/locus tag="RPA0633"
/function="InterPro IPR000100"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="probable ribonuclease p protein component
(protein c5)"
/protein_id="CAE26077.1"
/db_xref="GI:39647557"
/db_xref="GOA:Q6NC40"
/db_xref="InterPro:IPR000100"
/db_xref="UniProt/TREMBL:Q6NC40"
/translations="MSRLRQADPIAVAGGLRINGPAFLVLSLRREDQGPVIRVGFVT
KKTGTATERNIRRLRELVKRTDRGFLLPHSDYVLVGRRAALSRDPAFVWMLDLRTAL
ERVARSPPFA"
complement (2401..2535)
/genes="rpmH, L34"
/locus tag="RPA0634"
complement (2401..2535)
/genes="rpmH, L34"
/locus tag="RPA0634"
/function="InterPro IPR000271 COGs COG0230"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="possible ribosomal protein L34"
/protein_id="CAE26078.1"
/db_xref="GI:39647558"
/db_xref="GOA:Q6NC39"
/db_xref="InterPro:IPR000271"
/db_xref="UniProt/TREMBL:Q6NC39"
/translations="MRTYQPSKLVKRRRHGFARLATTGGRKVLARRARGKRLSA"
"
2866..4344
/locus tag="RPA0635"
2866..4344
/locus tag="RPA0635"
/function="InterPro
IPR003594:IPR003660:IPR003661:IPR004358:IPR005467 COGs
COG0642"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="sensor histidine kinase"
/protein_id="CAE26079.1"
/db_xref="GI:39647559"
/db_xref="GOA:Q6NC38"
/db_xref="InterPro:IPR003594"
/db_xref="InterPro:IPR003660"
/db_xref="InterPro:IPR003661"
/db_xref="InterPro:IPR004358"
/db_xref="InterPro:IPR005467"
/db_xref="UniProt/TREMBL:Q6NC38"
/translations="MOAADKOPETHAWRGLEPKIGLSKLLLTITPLIALILIV
PSIANYRLNRLDLAANTALVLDASGRVPSDLAQILDSIEARVAIKMGQOR
RLASNLPTKIDVDMNRVTPWGAILDSFVWLESNPAIRVVGPAENAEFIEV
IDAAPLUNMAYFSSNLLVSLTITSLGLIYLALHFLFVRPRLRTAMVNFRRDP
ESPASTVTPGPDDEITGLAERLSDMDRLVSLHQSRLAALGLAVSKINHDRLNL
ASSQLSDLSVPDRVQFAPKLVSLERLAFQCSLSYGRAQEAAPDRRLSILIE
PLVNDVRETAGLADEASITWVAIERGLTIDADPDOLFVLIANLVNAAQALEGQPRN
DANLOQIRITGRREGSVTVLESDTGPVQPKARDHLFEAFHGSVRAGSGGLGLAIAA
ELVRHAGGBINLVEGTIGATFRISIPDRPVDLHLSKSERAA"

```

```

misc_feature
order (2944..3012,3394..3462)
/locus tag="RPA0635"
/notes="2 probable transmembrane helices predicted
byTMHMM2.0"
4594..4670
/genes="trNA-Arg1"
/locus tag="RNA_7"
4594..4670
/genes="trNA-Arg1"
/product="trNA-Arg"
/evidence=not experimental
complement (4784..5293)
/locus tag="RPA0636"
complement (4784..5293)
/locus tag="RPA0636"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="CAE26080.1"
/db_xref="GI:39647560"
/db_xref="UniProt/TREMBL:Q6NC37"
/translations="MLKNLILFAANDHEHPALAAALREHNPGLTVSDDLLELCELE
PELQBSRLICFGPLTLQSDVVKQICHGAYQFYAPLQYPLGLAPSTAQETDDGPCV
SVIASIGDAAPFDQIIDITVTLPANIEAPQKDILAFTRLARFWRLSHAIACETNP
AEPSTRLN"
complement (5320..6648)
/locus tag="RPA0637"
complement (5320..6648)
/locus tag="RPA0637"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="CAE26081.1"
/db_xref="GI:39647561"
/db_xref="UniProt/TREMBL:Q6NC36"
/translations="MIPTSPSAKLTQLQAGSLRGLTSQRRRLDMLDRVGPVRLQRP
GEAPDRDAGLVITVWGGVAGAEFCNALPGCVVLPFGENPDPLKSLTFEGTV
GPCAGRGHEWVGGVDAETLVADAGETARPRIVSCYOHGCCDILVPLRQAASFC
AAHIGPLMSECDRLVAYEKAFILQAWELPEPLIYIENGALQYPLLPALQDCDV
ALHWRNLSARTLYFRGSDAAKALLGLRRRLAKAHPAVMDGLYLDQAWSLASLSA
LDTWPLRSYTLHGDPAPQSAVI VHDLPGLSADLPGDFASLVLTARRAGTGGRD
ALVMTSHQASGDGVAVILRTADTSRVVATTVETATSAFADNNGSGFGLMLLCAW
RDEVPFARAAAARLVAEIVFGQPIPHDFEARLDRQDAGAPSLIARG"
complement (6725..9127)
/locus tag="RPA0638"
complement (6725..9127)
/locus tag="RPA0638"
/function="InterPro IPR001951 COGs COG1011"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAE26082.1"
/db_xref="GI:39647562"
/db_xref="InterPro:IPR001951"
/db_xref="UniProt/TREMBL:Q6NC35"
/translations="MSFVSDIHTPRAQNRIRMSFDVDTFLIRACTTPTGVPFELTVQL
SRVSLDPCNMSEHFVQHRILAESRVREASRGATEINIDQIYARFPFLFGLERED
LNHLIEAFSABELCIRNDMLQEQYRARRRAGNVGFI SDTYWNTWRLGLLRACSP
GLTWFLYASCDHSGSKGDELFATYLRQQGVASSAYHVGDNHADIIRGAKRHGIRP
HYQAGPRLTTLRLLEDLLQQLMFKGPQPLDRGARTLRMVAARSAEQAASHLGST
VYGPVLAADQFVATERADLAGGRVAIGFLRGDGLFSLRLWQQSGRGPAAVLEINR
RVSILASADTIGPICELIRKIPRIDAKLQDMKILPPALADFPACDCGIATGABELA
DALPNLIEPHEITAMLAAGURGLDITRQOIPDLDCTDILLVLDLGSVQALSRV"

```

Query Match 3.9%; Score 76.8; DB 1; Length 349260;  
 Best Local Similarity 50.6%; Read, No. 1.5;  
 Matches 239; Conservative 0; Mismatches 227; Indels 6; Gaps 2;

QY 819 CTTCTGTACACGCACTCTGTGGAGAGACCCCGAGCCGATATGGAGGTATGGAGATCA 878



LOCUS AY386263 137241 bp DNA linear VRL 20-JAN-2004  
 DEFINITION Orf virus strain OV-IA82, complete genome.  
 ACCESSION AY386263  
 VERSION AY386263.1 GI:40019122  
 KEYWORDS  
 SOURCE Orf virus  
 ORGANISM Orf virus  
 Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae; Parapoxvirus.  
 1 (bases 1 to 137241)  
 Delhon,G., Tulman,E.R., Afonso,C.L., Lu,Z., de la Concha-Bermejillo,A., Lehmkuhl,H.D., Piccone,M.E., Kutish,G.F. and Rock,D.L.  
 Genomes of the Parapoxviruses Orf virus and Bovine Papular Stomatitis Virus  
 J. Virol. 78 (1), 168-177 (2004)  
 2 (bases 1 to 137241)  
 Kutish,G.F.  
 Direct Submission  
 Submitted (08-SEP-2003) African Swine Fever Research, Plum Island Animal Disease Center, U.S. Department of Agriculture, Agricultural Research Service, P.O. Box 848, Greenport, NY 11944-0848, USA  
 Location/Qualifiers  
 1. 137241  
 /organism="Orf virus"  
 /mol\_type="Genomic DNA"  
 /strain="OV-IA82"  
 /isolate="ORPA"  
 /isolation\_source="nasal secretions of a lamb at the Iowa Ram test Station during an outbreak in 1982, then passaged in ovine fetal turbinate cells"  
 /db\_xref="taxon:10258"  
 /country="USA: Iowa"  
 complement (2409..2858)  
 /codon\_start=1  
 /evidence=not experimental  
 /product="ORF001 hypothetical protein"  
 /protein\_id="AAR98099.1"  
 /db\_xref="GI:41018491"  
 /translation="MHPSRRLLGALLVALGFLGLFRPAAPPLPAALVAGPVRA NGASVTCITVGDCRHMVAHGGTLPVPLAGMHTASLRKGLLLNVATVH IYDVRLEAPFELTCVAVAGGYNAAWATRAEWRRLQAQHRSEL"  
 complement (3002..3355)  
 /codon\_start=1  
 /evidence=not experimental  
 /product="ORF002 hypothetical protein"  
 /protein\_id="AAR98100.1"  
 /db\_xref="GI:41018492"  
 /translation="MTPTSRSELLTLLVHLVRAARRPSPALVRLAAGAAAGPDA AGRPLAALLRALRRGGPLDARLLPLLARRGAAASARARGGARSAAGLLLAAGARW RLALAAAGRAAG"  
 complement (3713..3928)  
 /codon\_start=1  
 /evidence=not experimental  
 /product="ORF005 hypothetical protein"  
 /protein\_id="AAR98101.1"  
 /db\_xref="GI:41018493"  
 /translation="MPGEGVHYSHSGTSGVASTPGGAVKRPSSGGGIGPLEKRM CVGGGGGSKSISGGVNGVNGGLGKI"  
 complement (4012..4491)  
 /note="similar to Vaccinia virus strain Copenhagen F2L"  
 /codon\_start=1  
 /product="ORF007 dUTPase"  
 /protein\_id="AAR98102.1"  
 /db\_xref="GI:41018494"  
 /translation="MEFCHTETQVVLRSQNTIPARGSPGAAGLDLCSAHDVPIPSH CSRNVTDLLIPSPSCYRIAPRSLGAVKHFDVAGVIDEDYRGNVGVLNFGNS DFEVKGDRILICERISCPAQVEVCLNDTRDGSFGTGLACGCRDTAWTIS"  
 complement (4554..6104)  
 /note="similar to Vaccinia virus strain Copenhagen B4R"  
 /codon\_start=1  
 /product="ORF008 ankyrin repeat protein"

/protein\_id="AAR98103.1"  
 /db\_xref="GI:41018495"  
 /translation="MLSRSEVVPHADLLFRYLESGQVLETVRALVATDADVNFRRGE HGRTPLHLCHFARHEQCAIYRVLLEAGADYAKDSCGRTPHLHVQDVGPRPVVA LMLEAGADVCDASVVVCDVSLFACPGDGAELVARLLLDAGRVNEGTYGVT PLHVAKOWIREDLRLLLRGANPCDCLGTPLAALLCGSGVSAALVDMMLRAG ADACVVDAVGRITLHLARTAKISGMVRLTGLVPAADACGNTLMHLYATGSC ERGVDFVLRLGLDLNLRNNLQTHALRAVESHGACCLVRMGAELEHMSASGLCAV SEMLRNNVRATAAVALRPPTELLARALITSEQMGHFFRSEALLCQVELLRGEG ARMLAARALADYATVIRACEBIASMRVRCHTDALTLDVLRADDAKALFVQNAFLE RRAAFPIYGTALFGKICMRLRVSLAEQIAGLMCPALPPEI VTSILCFLSYDILLDL RRAMLTRP"  
 complement (6265..7626)  
 /note="similar to Vaccinia virus strain Copenhagen F11L and Molluscum contagiosum virus MC018L"  
 /codon\_start=1  
 /evidence=not experimental  
 /product="ORF009 hypothetical protein"  
 /protein\_id="AAR98104.1"  
 /db\_xref="GI:41018496"  
 /translation="MFFWFCGRPRMSALARLFGRRGRNOCVPTAGCWPLRRPRDG ADGGSDDDYGYGDTDSVASSVAPVEARPLVLTVPGTCPVLVDSVFTRIENTSVI VTHGSEWTDISLFESTFHRAPARICRSESRVVRDANMHFEVSGHGLKHKHDEWVR HMAPTVAAYVTIICVNEGMAIGISVNTKELKTMQEGDLVVPARSVFLPQVGGI DAYIILVLPVTAELCDLGFVPPSIDQEAQVRVANVAETRAKACKIVNQIVKLAVD LEKCYEICNLVIMITEFNAYNDVGAQMTASVDIAIARGMGAATQALCRQSLRRR PLKSAALTRNHSILHEGITSFVAFQTPGTPHPDLSMARMERDLELYSKFSELWDD ILQLASDLSDLTLPCECIERYIHHMVCSEAGVGRNELVQRLTMIAGAYRLTSGSGG VPV"  
 complement (7651..9582)  
 /note="similar to Vaccinia virus strain Copenhagen F12L and Molluscum contagiosum virus MC019L"  
 /codon\_start=1  
 /product="ORF010 putative EEV maturation protein"  
 /protein\_id="AAR98105.1"  
 /db\_xref="GI:41018497"  
 /translation="MLGFWGKCGVDPDQANDEPPPTDLVPGEYALAKARDGKTMIF DAEGARAADVLPKLSVRISCTFIDRADALAPPAPATQRFATVSRHDDPCYIFDS SSPFLDIILQRAETPELRAEFDDPPRDLASMDLQNLSDVGLDLMVSHDAVR RIGTLRQLTLLDMNVCCVGYAIWIKDPSTYSRPTDEVLSHDVTRTASASWAQPH VHAPGNSVAAVVFVFGSDTVRLNSIATNPAGVPVPGCAKLVRDPLMWLTCKEN ITVGYFSFPDTRQVMELAGKWTISENTLLSLRGNRVAHQALARPATLFSYVC EFWGGVVDVDPDVVDESASAAAASAAAQALSDAALAHQALARPFPSCDMAAF RGLREMLVNAARGCAACPVHFSALNVEAALFLEETGASVPFADARCFRLAYLREV ASQLYPGKPRVYVSKLITRGLLAALCEVTRSAVKIPVLPFPADESLRFSVLTSDV IETALRLGYEIRVVCALGEGRCDEVLKAGLEAQMTRATLRLNIPTSLMSRTASMPLE ALETDSDSPGRCASVVRAPAAVARRAVHGFIERVDCHATAAFVLRHGYDRFWVRERA ARCLAGVSGIQEVLPSVR"  
 complement (9653..10789)  
 /note="similar to Vaccinia virus strain Copenhagen F13L and Molluscum contagiosum virus MC021L"  
 /codon\_start=1  
 /product="ORF011 putative EEV envelope phospholipase"  
 /protein\_id="AAR98106.1"  
 /db\_xref="GI:41018498"  
 /translation="MWPFSIPLGADCRVETLPAPVAVSLAQNMSTLPCFTAITSSA KFLYICSPCCNLSTKGGVDKGLCTLAKEGVNVTLVVDQSKDADLEAGVNV YKVKVSTREGVNLGSLFWSLDAGHWYVGSASLTGSGSVSTIKNLGLYSTNKKHLAWDL MRYNTFVSMIVEPKVPFTRLCVAVVTPATNFHLDHSGGVFFSPSPERFLGFYRTL MRYLVHRIENAKNSIDLSLSNVPVIKHASAVVEWPRIIDALLRAIRDGVRVRLI TEWKVADPLVSAARSLDDFGVSDMSVRKFVPGRDDAANAKULLI VDDTTPAHUTV ANLDGTRHYTHAFVNAEKGDIVKDLSAVFERDWRSEFCFPI"  
 complement (10818..11087)  
 /codon\_start=1  
 /evidence=not experimental  
 /product="ORF012 hypothetical protein"  
 /protein\_id="AAR98107.1"  
 /db\_xref="GI:41018499"  
 /translation="MLIYGPLREVEDLIDRMDDVQVSGDLMPPTDPVPEDVAQCSPE DMANMLDGLIEDIQLKSELQLTSEMEBELLREBELARLLKDFP"  
 11396..11635  
 /codon\_start=1  
 /evidence=not experimental

CDS

CDS

CDS

CDS

CDS

```
/product="ORF013 hypothetical protein"
/protein_id="AAR98108.1"
/db_xref="GI:41018500"
/translations="MGVWMCGLFVWAGWIAARRVFCVALRORVSRDKGYVAVIOTC
DDDFTEEFDDQVALLRDRVSRVVAAPAGTE"
complement (11671..11952)
/notes="similar to Molluscum contagiosum virus MC026L"
/codon_start=1
/product="ORF014 modified RING finger protein"
/protein_id="AAR98109.1"
/db_xref="GI:41018501"
/translations="MPVKVKQRLAVSCTILACSDVCYIQNKRSPGCIINACBGPFP
QVWVGCHVHCITPTNTEVCFVRSALQKAEPLDMSTSKSAREFIV"
complement (11955..13574)
/notes="similar to Molluscum contagiosum virus MC027L"
/codon_start=1
/evidence=not experimental
/product="ORF015 hypothetical protein"
/protein_id="AAR98110.1"
/db_xref="GI:41018502"
/translations="MEELRTRLIAQMEMDLSVAHISRLRLTLKRVNTFTAVACLR
EALARKSWPPGRLGTDIHEDMTVLEYVMQVEVENARENIEPYLCVIOAHNADVLDD
CLREADMEYGLRIGRVCREFPHLVADRCSCLSAFSEDA CVRFTNLMRQS
RTMARCIVRAHSGSYDELVDVMDVRHQNVRVWVNVANVEDVIAADLEGLMAY
CRTCLMSARVASGELIDLKDRMPLIVECIETHVAESNLDRALCLVNVYVDLKTGH
LHARLERSGKALYGNVAAAPCDTMDLLSYLERLPRELAARFARVAFARIVKSVYT
GVNLEKLSUKAWELPPNASSALNDLVRLMSGRLLPLLPPOPMEHDEGESHH
DPMVLMPNVVYTFVTPPLAAMTPPESLRGAVESALARADTAYDVAPLAAGRAELT
IPAAGREQEVI CSSAHLVCLVIMDEAGKTGTVTEDAARHLGNDARLAEMTLTSLCRE
KIAKRTVDGVRPALNPLEARDSPWVFN"
complement (13644..14423)
/notes="similar to Vaccinia virus strain Copenhagen F16L
and Molluscum contagiosum virus MC029L"
/codon_start=1
/evidence=not experimental
/product="ORF016 hypothetical protein"
/protein_id="AAR98111.1"
/db_xref="GI:41018503"
/translations="MRIDGMEVEVDENATRAARGMILLSTFMSWMDIVRQBELA
RQCYKNGMTPCVSVTRIGRIDASNGEDTWEARSHDVRHVFLFKASVVDVHVDLFL
VGLGELGLDLYFVADGFPVDSMPRFPVDMETMLRCLVRDYMEDVGGGLHPR
EPGSAQCLMDRVMYFPGLLAMOAHFARRSRRAPTPLPRASPRPOTRRSPRUGVAD
SANYQALDARANNQLRRRRDDVLSRLNLDNRVDD"
14677..14994
/notes="similar to Vaccinia virus strain Copenhagen F17R
and Molluscum contagiosum virus MC030R"
/codon_start=1
/product="ORF017 DNA-binding phosphoprotein"
/protein_id="AAR98112.1"
/db_xref="GI:41018504"
/translations="MDGVHRRPPIVNDGKGVILVRLRMCEVPEAKCEGRASCV
LKMDDPRSPSCERRPLSPSPCPMRTPPGSPLOAPLMRTQMLQGLFDAKNGEOMC
RRQ"
complement (15018..16436)

Query Match      3.98; Score 75.6; DB 13; Length 137241;
Best Local Similarity 43.1%; Pred. No. 2.6;
Matches 463; Conservative 0; Mismatches 609; Indels 2; Gaps 2;

QY 618 CGACCAATTCATCGGCGCCGAGCGCGCGCTACCTGAGCAGCAAGCTGGAGCGCGT 677
DB 20215 CGAGTCTTCGACGCCGTCTCAAGCGCGCTTCGACGCGCGCAGCTCTTTGCGGGC 20274

QY 678 GTGGAGCAGAACACCAGGGTTTCGATCCCTACGTGCGCGTGGCGCCCCCTACTA 737
DB 20275 GCTTCGCGGCACCTTCGAGGCTTCGTTGCGGCATCCGGGCTTACCGCGCAGCTTCGC 20334

QY 738 CGTGTGATTTGGCGGCTGCCAGCGTTGGCCACGCCCTTCACGAGGAGCGCGTGGAGCG 797
DB 20335 GGAATCTGCGCGACCGCGCTTCATGGAGGTGGCGCGCGCGCGCGAGCTTCGCGGA 20394

QY 798 GCGCCCATGTTCCCGCCACCTTCCTCTGTACACGACGTGCTGGAGGACCCCGAGCCGGA 857
```

```
DB 20395 GGTCAATTGGCTGAACCAACGACATCGCCGCAACCCGCTCTTCGCGAGCGGAGCCGCT 20454
QY 858 TATGGAGGTGATGGAGATCAACCCCAAGGACACGGTGTGCTGACCTGACTAGCCGCGGCTG 917
DB 20455 GCGCGACGCGGAGCTCATCTTTCGAAAGACCTTCCGCAAGACCGAGTTCGCGCCGCTCAA 20514
QY 918 CAATGCCCTGAACCTCTGCTGTCAGGGGCGCGCCAGGTGGTGTGCTGCTGAGCTGCAACCC 977
DB 20515 GCGCTCTCGCGTCTCGGGCTGCTGTCTGGGCTTCCTCGTGAAGAGGACCTTGGCGG 20574
QY 978 CGCGCAGTCCGGCGCTTCTGGAGCTGAAGAGGTGGCCATTCAGCAGCTGGAGTTTGAGGA 1037
DB 20575 CGAGTACGCGGCAACACGACCGCAGGACCTGTTTACGCTGCTCAGAAGGCGCGGCGCC 20634
QY 1038 CGTGTGGAGCTGTTTGGCGAGGGCTGCACCGCGCATTTGAGAGCTGTACGAGAAGAA 1097
DB 20635 CGTGGCCCA- CAGCGCGCTCACGAGAGCATTCGCGAGTACCTCTTCCCGCGAGACAGGC 20693
QY 1098 GCTGGCGCCCTTCTGCTGCAACCA- GCCACAACCTTCTGGTCCAAAGCGCTCTGTGTACT 1156
DB 20694 CAGCACACTGGGTCTGGCTGACCGCGCGTGGCCGACGCGGAGGTGTACCGGACC 20753
QY 1157 TCAGACACCGCTGTACTACAGGGGCGCATGGGCAAGTGTGCTGGGTGCTGAGTGCC 1216
DB 20754 GGCCCGCGCGCAGCTTACAGAGCGGTGCTCAGCTACGCGTACTCAGAGGTCAAGCAGG 20813
QY 1217 TGGCCCTGTGTGGGACTGGGCAAGACCGTCAAGCGCTCGCCCAAGCGCCACATGG 1276
DB 20814 GCGCGCTGAACGCGCAACACGCTCGTGTACCGCGTTCGAGGACGACCCGACATCA 20873
QY 1277 AGGACAGCGCGCTCTGTGGGACAGCAACATGTCTATCCACTTCGTGAAGAACGGGCCCA 1336
DB 20874 AGGTCTGTCTGCTGAGCTCATCTAGCGTGC CGGGACATCTCGGCGTCTGTGACT 20933
QY 1337 AGCGCTGTGTGGCTTTCGTCAA GTTCTGTGAGCTGTGTCTTCAA CAAGCGCGTGC 1396
DB 20934 CCGCAACAGGAGTGGCGGAGCTACTTCTGTAGTCTGTACCGCGAGAACTTCGTGACG 20993
QY 1397 TGTGTTTCGCGCGCGGTGCGGCAAGCAGTACCGCTGTATCAAGCGCGGACGCGCATCC 1456
DB 20994 GACGCACTTTCACCTCGGACGCGCGTTCGCGACGACCTCTTTCGCGGTGGTCCGCCCG 21053
QY 1457 CCATTGAGAACTACATCGCGCGCACCATGGAAGCGCGTGGCGGAGAACTCCGACGTGCGCA 1516
DB 21054 TCAGCGCGACTTCTTCGAGCGCGGCGCATCCGCGAGGCTTCAGTGCAGACGCGGCG 21113
QY 1517 AGCAGAACTACTTCTACTACACTGCTCACCGGCAAGTTCTCTGGCGGACAACTGCGCCCA 1576
DB 21114 TGGAGAGCGCTTCACGACATGGACTCAACAAACGCTTCATGTCTGACCTCATCTACG 21173
QY 1577 CTTACTTGGCGAGGCGGCTTCGCCACCTCTCAAGAGTGGCGTGGTGGACAACTGACCG 1636
DB 21174 ACTCCGTGAGCCCCGAGCTCGCGCGCGCGAGCGCGGCTTCGCGCTGCGGTGCACAACG 21233
QY 1637 TCTCCACCACTTCTTCATGGAGGAGCTCAAAGCGCGCACTACACCAAGGTGA 1690
DB 21234 AGACTCCGACTACTTCATCCGAGGATCAACACCTACCTCTTCTCTCAGCGAGA 21287

RESULT 13
AP006840_19/c
WPCOMMENT
Sequence split into 36 fragments LOCUS AP006840 Accession AP006840
Fragment Name      Begin      End
AP006840_00        1          110000
AP006840_01       100001      210000
AP006840_02       200001      310000
AP006840_03       300001      410000
AP006840_04       400001      510000
AP006840_05       500001      610000
AP006840_06       600001      710000
AP006840_07       700001      810000
AP006840_08       800001      910000
```



AP006840\_09 900001 1010000  
AP006840\_10 1000001 1110000  
AP006840\_11 1100001 1210000  
AP006840\_12 1200001 1310000  
AP006840\_13 1300001 1410000  
AP006840\_14 1400001 1510000  
AP006840\_15 1500001 1610000  
AP006840\_16 1600001 1710000  
AP006840\_17 1700001 1810000  
AP006840\_18 1800001 1910000  
AP006840\_19 1900001 2010000  
AP006840\_20 2000001 2110000  
AP006840\_21 2100001 2210000  
AP006840\_22 2200001 2310000  
AP006840\_23 2300001 2410000  
AP006840\_24 2400001 2510000  
AP006840\_25 2500001 2610000  
AP006840\_26 2600001 2710000  
AP006840\_27 2700001 2810000  
AP006840\_28 2800001 2910000  
AP006840\_29 2900001 3010000  
AP006840\_30 3000001 3110000  
AP006840\_31 3100001 3210000  
AP006840\_32 3200001 3310000  
AP006840\_33 3300001 3410000  
AP006840\_34 3400001 3510000  
AP006840\_35 3500001 3565135

Continuation (20 of 36) of AP006840 from base 1900001 (AP006840 Symbiobacterium thermophilus)

Query Match 3.8%; Score 74; DB 1; Length 110000;  
Best Local Similarity 44.5%; Pred. No. 4.3;  
Matches 293; Conservative 0; Mismatches 365; Indels 0; Gaps 0;

QY 658 GAGCAGAGCTGAGCGCGTGGGAGCAGACACACAGGTTGCGATCCCTACGTGCGG 717  
DB 53945 GAGGCCAGAGCACTGCGCGGGTGAAGCAGGCGGGGAGCAGCAGCGTGGAG 53886

QY 718 TGCTCGCGCCCCCTACTACGTGTGGATTGGCGCGCTGCCAGCGTTGGCCACGCGCTG 777  
DB 53885 CGGCTCGCGCGCAGATCGACAGCTCCGCGAGCGTGGCGCCCGCAGCAGCGCGTG 53826

QY 778 CAGCAGAGCGCTGAGCGCGCGCCCATGTTCCGCCCACTTCCTGTACAGCAGTCG 837  
DB 53825 GAGGAGCGCGGAGAACCTTGGCGCGCGGAGCGCGCTTACCTGCCAACCCCTCCGAC 53766

QY 838 TGGGAGGACCCCGAGCGCGATATGAGAGTGATGAGATCAACCCCAAGCAGCAGCGTGTG 897  
DB 53765 TCCAGCGCGCGAGCGGTATCGCGGCTGAGAGCGCTTACCGCGAGCGGTGCCAAG 53706

QY 898 ACCCTGACTAGCGCGCGCTGCAATGCGCTGAACTTGTGTGTCAGGGGCGCGCCAGGTG 957  
DB 53705 AGCCAGCGCAGCGTGCAGGAGCGCGCGAGAGCTCGCGAGCGCGGAGCTGGCC 53646

QY 958 GTGTGCGTGTGATGCAACCCCGCGAGTCGGCGCTTCTGAGAGTGAAGAGTGGCCATT 1017  
DB 53645 GCGCTGTGAGGCGGAGCAGCAGCGCGGTAGCGCGAGTTGAACGTGCGGCTGGGAG 53586

QY 1018 CAGCAGCTGAGTTTTCAGGACGTGTGCGAGCTGTTCGGCGAGGCGTGCACCGCGCAT 1077  
DB 53585 ATCCAGTGTGCGCGGAGTGGCGGTGAAGAGCGCGGAGCGGTGCGCCCGCGG 53526

QY 1078 GAGGAGCTGTACGAGAGAGCTGGCGCGCTTCTGTGCGCAACACGACCACTTCTGTG 1137  
DB 53525 CCCGACATGGCCAGGTGGAGCTGGCGCGCGCGCGCGGTGAGCGCGCGGACCGCTG 53466

QY 1138 TCCAAGCGCTTGTGATCTTCCAGCAGCGGCTGTATCTACAGGCGCGGATGGCAGAGCTG 1197  
DB 53465 GACAAATGCGCGCAATCTGAGGCGGACCCGATCGTCCCGCGCAGCGCGGAGCGTG 53406

QY 1198 TGCTGCGTGTGAGTGCCTTGGCGCGTGTGCTGGGACTGGGCAAGACCGTCAAGCGCCTC 1257  
DB 53405 CTGAGCGTGCAGTGAAGCAGCGGACCCCGGTGCAGCAGGCGCTTCCCTGTTGCGCGTC 53346

QY 1258 GCCAAGCGCCCAATGAGGAGGAGCGCGCTCTGTGGCAGCAGCAATGCTCATCC 1315  
DB 53345 GCGCAGATGAGCGGATGGAGCTGATCGCCAGGTGAGCGCGCTGAGACGCC 53288

RESULT 14  
BX842626 75337 bp DNA linear PLN 17-APR-2005

LOCUS Neurospora crassa DNA linkage group I BAC clone B18P24.  
DEFINITION BX842626  
ACCESSION BX842626  
VERSION BX842626.1 GI:38566958  
KEYWORDS  
SOURCE Neurospora crassa  
ORGANISM Neurospora crassa  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.

REFERENCE 1  
Schulte, U., Aign, V., Hoheisel, J., Brandt, P., Fartmann, B.,  
Holland, R., Nyakatura, G., Mewes, H.W. and Mannhaupt, G.  
Unpublished

JOURNAL 2 (bases 1 to 75337)  
German Neurospora genome project.  
AUTHORS Direct Submission  
TITLE Submitted (26-NOV-2003) MIPS, Institut fuer Bioinformatik,  
GSP-Forschungszentrum fuer Umwelt und Gesundheit, GmbH,  
Ingolstaedter Landstrasse 1, D-85764 Neuherberg, FRG, E-mail:  
G.Mannhaupt@gf.de Project Coordinator: Ulrich Schulte, Institute  
of Biochemistry, Heinrich-Heine-University, D-40225 Duesseldorf,  
E-mail: ulrich-schulte@uni-duesseldorf.de

COMMENT BAC clone 18P24 (strain OR74A) is available at the Fungal Genetic  
Stock Center, <http://www.fgsc.net>  
Sequencing was performed by MWG Biotech AG, Ebersberg, Germany,  
<http://www.mwgdna.com>  
Information on performance of analysis and a more detailed  
annotation of this entry and other sequences can be viewed at:  
<http://mips.gsf.de/proj/neurospora>.

FEATURES  
Location/Qualifiers  
1..75337  
/organism="Neurospora crassa"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:5141"  
/chromosome="I"  
2973..3008  
/note="36 bp gaaagaagaagaagaaga tandem repeat"  
4630..5673  
/gene="B18P24.020"  
complement(4630..5673)  
/gene="B18P24.020"  
complement(4630..5673)  
/note="B18P24.020"  
BG279679, BG279920, BG278902, BG278911, BG279791"  
/codon\_start=1  
/product="putative protein"  
/protein\_id="CAE76260.1"  
/db\_xref="GI:38566959"  
/db\_xref="UniProt/TREMBL:Q7SDH3"  
/translation="MSTKPIKLSTPTVTFANFPNIPRPTDELRSASVLPALSRLPLS  
ATRSAPRDLPSALPSAGLPSPGFSATLKVDHDLQKTPITPPVAYLDFLKMQLVS  
PSLAPPTQKGYLHRTSTASSLSTSTSNHSESSVDSATTDSEKEGQIESAPS  
TARTDVSSEPEIKVEAEKEDKKVDEKEDKEDENVDNGKEDKSRPAITISQPTTP  
QSSLYPMGAPATGAAVFPSTKIPASPAISTSAISYSPRTPLSASVSPDFDWTALKRI  
RFABYSSLSKRSAPSDAPAVSTPTSPVPAKESRSIRHIREVTVTRVTPRMAPAPKG  
KKRKIDADTAIKS"  
complement(4630..5673)  
/number=1  
7464..7512  
/note="49 bp poly(a) tandem repeat"  
14161..15111  
/gene="B18P24.030"  
complement(join(14161..14342,14412..14477,14566..14717,  
14924..15111))

```
/gene="B18P24.030"
complement(join(14161..14342,14412..14477,14566..14717,
14924..15111))
/gene="B18P24.030"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAE76261.1"
/db_xref="GI:38566960"
/db_xref="UniProt/TREMBL:Q7SDH4"
/translation="MARSAAPGPGQAEARPHREFPQPPPAQWKRCGPOOTSET
RNDPAGLPGNHAPVAGSTRGDDAVLAKDAAQSGIIVSGRCPLGVFVPCAGN
ESDMWNEAVRYNLPIARVKVGLDCWLSVRDPAVLDSHPCPVPLALGFTGLRCD
GGGSRWGTMLVLRRLRAAALRMSLSVSV"
complement(14161..14342)
/gene="B18P24.030"
/number=1
complement(14343..14411)
/gene="B18P24.030"
/number=1
complement(14412..14477)
/gene="B18P24.030"
/number=2
complement(14478..14565)
/gene="B18P24.030"
/number=2
complement(14566..14717)
/gene="B18P24.030"
/number=3
complement(14718..14923)
/gene="B18P24.030"
/number=3
complement(14924..15111)
/gene="B18P24.030"
/number=4
1595..17623
/notes="29 bp poly(g) tandem repeat"
19883..20827
/gene="B18P24.040"
19883..20827
/gene="B18P24.040"
/notes="similarity to HAPB protein, Aspergillus nidulans,
TREMBL:ENHAPB_1
contains EST gb:AA898026, AA898402"
/codon_start=1
/product="related to CCAAT-binding factor HAPB protein"
/protein_id="CAE76262.1"
/db_xref="GI:38566961"
/db_xref="GOA:Q7SGY4"
/db_xref="InterPro:IPR001289"
/db_xref="UniProt/TREMBL:Q7SGY4"
/translation="WMEYACYQPAQSHSNPHMTAYSSSAGGNSITSPSSOHLPTS
PILOTQHQSPSTQGHSMYQTVGVQONHGYCMPLQAAALANTAASGTNYNPMYA
ADNLQSPRMSGVNSKQVRAGRSPFPQNSIPPQORLSQVSPGVNQAQWLNHA
SRVAPPPMTAQQMPQPPQSPENASGAVESPLYNAKQPHRLKRVARQKLEQRL
LTNKRKPYLHESGRHNHMRPRGPGRFLTADAVQMRDKVNGVDKQDGSQSSVT
AGSKTTGTRKRAESTSGAPNKKAKAEPSPEDDADSD"
19883..20827
/gene="B18P24.040"
/number=1
21444..21471
/notes="28 bp gcacgcacgcacgcac tandem repeat"
23226..26083
/gene="B18P24.050"
join(23226..23780,23851..24177,24239..26083)
/notes="no similarity to N-terminal half of the protein
similarity to S-adenosylmethionine:diacylglycerol
3-amino-3-carboxypropyl transferase btaA, Rhodobacter
sphaeroides, TREMBL:AF329857.1
contains EST gb:BF072578, BF072579"
/codon_start=1
/product="related to S-adenosylmethionine:diacylglycerol
3-amino-3-carboxypropyl transferase btaA"
```

```
/protein_id="CAE76263.1"
/db_xref="GI:38566962"
/db_xref="GOA:Q7SGY5"
/db_xref="InterPro:IPR000051"
/db_xref="UniProt/TREMBL:Q7SGY5"
/translation="MGDNSAMASHGHMGNIISLPSLSTVTRTLADLNFNNTTYSIFFT
GVAVLFLVTSNYKRTTKNEDNEGNPSLSKSLLLFCYSCFIKPHATAGTTGTQ
QDALSFVRSQADIYDATRTLLKGRDMLAASQRLKYKVEAGLGGAGDGLKRR
QNGKTCVAGTGTGRKPIWVDVGGGTGWNIEAMAKFVNVSEFFKTVLVDFSPS
LCEVARKFARLGWENVRICTDARKFERLVEDVDGSGSDSPSSGMMGETKP
GRHAGNELTMSYSLSMDDPYSLLDSLESLLAPHLIAVDVFAQSKVDFTPRNTG
GLNCRHVGTFARNFWRFDADKRLSPLEPARRPYLEYRGTVIIVARNNTILGAIPYI
WGLCKKPTSSLSLPHIEIHDAITAEPRSPRSLVGRHSSATNALAFAGVGTAT
MRSAKNTAIEINISANPLPFFYQNHWRYYDDQLPHTQFNDEYIYAFWEDSRV
DRELLNLGDDVLAITSAGDNLISYLMQSPARVAIDLNPAQNHLLEKVASFTLID
YDVWVKIFGEGKHPDPSLLISKLSPHSGRAFOYLSNAHIFTDPAGRLVDTGGR
YAIRFRWISTLFFCSEAVRLLSTPTLEGQSIYHTKIRPCLLNFVGLVLSDAF
LWSALGVPNQVAMEADYHRSSISSTTPSSKEKPSRAELIHYTTSHLDPVLSH
LSDNPFYLVGVLTQTPCHPDYLSAAHSILSAPGAFGLRIHTDEIOEVLARFQ
GTLITVAVMDSMDFDPPSPBEKEGRKAREQVRRLNALKVGKVLRSAGVEPWY
VRFVVEEGFGARRVCGRESGRGDQECIDRVNMYVASCWILEKMEDELVDASA"
23226..23780
/gene="B18P24.050"
/number=1
23781..23850
/gene="B18P24.050"
/number=1
23851..24177
/gene="B18P24.050"
/number=2
24178..24238
/gene="B18P24.050"
/number=2
24239..26083
/gene="B18P24.050"
/number=3
27766..28499
/gene="B18P24.060"
complement(join(27766..28118,28258..28275,28370..28499))
/gene="B18P24.060"
complement(join(27766..28118,28258..28275,28370..28499))
/notes="match of 21 ESTs
similarity to succinate dehydrogenase precursor (EC
1.3.99.1), YDR178W, Saccharomyces cerevisiae, PIR:A54380
contains EST gb:A1397618, AW18559"
/codon_start=1
/product="related to succinate dehydrogenase precursor"
/protein_id="CAE76264.1"
/db_xref="GI:38566963"
/db_xref="GOA:Q7SGY6"
/db_xref="InterPro:IPR007992"
/db_xref="UniProt/Swiss-Prot:Q7SGY6"
/translation="MASVARSSALLKQVAAQSSVAANGRLVRAAFHTTSKSLPPPPQ
RIEHTVNDVEVPVPPSPSHGSHWTDFRVAAGLILTVAPFAAGSLNPTMDVLAAT
ILHSHFTGNIIVDVPKRVKARKVETWGLNAATVLVGLALYEFETDVLGTETI
KRVWKA"
complement(27766..28118)
/gene="B18P24.060"
/number=1
complement(28119..28257)
Query Match 3.7%; Score 73; DB 15; Length 75337;
Best Local Similarity 51.0%; Pred. No. 6.2; Indels 0; Gaps 0;
Matches 172; Conservative 0; Mismatches 165;
QY 779 ACCGAGGCGCGTGGAGCGCGCCCATGTTCGCCGCCACCTTCCTGTACACGAGTCGT 838
|||||
DB 24804 ACCGAGTCAACTCCCGAAGCACACCCAGTTCATGACGAGTACATCTACGCCCTTTACCT 24863
|||||
QY 839 GCGAGGACCCCGAGCGCGGATATGAGGTGATGAGATCAACCCAGGACACCGTGTGTA 898
|||||
DB 24864 GCGAAGACTCGCGCTGCGAGAGAGAACTCTTAACCTCGGGCGCGAGCGTGTCTAG 24923
|||||
```

QY 899 CCTGACTAGCGCGGTGCAATGCCCTGAACTCTGTGTCAGAGGCGCGCCAGGTGG 958  
D 24924 CCATCAGCGCGCGGCAACAATTCTTCTTACCTGATGTCAGAGTCCCGCTCGGTGC 24983  
QY 959 TGTGCGTGGAGTCAACCCCGCGGTCGCGGTCTTCTGAGCTGCAAGAGTGGCCATTTC 1018  
D 24984 ACGCCATCGACCTAAACCCAGCCCAAAACCACTGCTTGAACCTCAAGTCCGCTCTTTTA 25043  
QY 1019 AGCAGCTGGAGTTTGAAGAGCTGTGGCAGCTGTTCGGCGAGGCGGTGCAACCCGCGCATTC 1078  
D 25044 CGACTCTGGATTACCCGCGCTCTGGAGATCTTGGTGAAGGCGCAACACCCGCACTTTC 25103  
QY 1079 AGGAGCTGTACGAAAGAGCTGGCGCCCTTCTGTCTC 1115  
D 25104 GCTCACTGCTCATCTCCAACTCTCCCTCACCTCTC 25140

## RESULT 15

AP006840\_24/c

## WPCOMMENT

Sequence split into 36 fragments LOCUS AP006840 Accession AP006840

Fragment Name	Begin	End
AP006840_00	1	110000
AP006840_01	100001	210000
AP006840_02	200001	310000
AP006840_03	300001	410000
AP006840_04	400001	510000
AP006840_05	500001	610000
AP006840_06	600001	710000
AP006840_07	700001	810000
AP006840_08	800001	910000
AP006840_09	900001	1010000
AP006840_10	1000001	1110000
AP006840_11	1100001	1210000
AP006840_12	1200001	1310000
AP006840_13	1300001	1410000
AP006840_14	1400001	1510000
AP006840_15	1500001	1610000
AP006840_16	1600001	1710000
AP006840_17	1700001	1810000
AP006840_18	1800001	1910000
AP006840_19	1900001	2010000
AP006840_20	2000001	2110000
AP006840_21	2100001	2210000
AP006840_22	2200001	2310000
AP006840_23	2300001	2410000
AP006840_24	2400001	2510000
AP006840_25	2500001	2610000
AP006840_26	2600001	2710000
AP006840_27	2700001	2810000
AP006840_28	2800001	2910000
AP006840_29	2900001	3010000
AP006840_30	3000001	3110000
AP006840_31	3100001	3210000
AP006840_32	3200001	3310000
AP006840_33	3300001	3410000
AP006840_34	3400001	3510000
AP006840_35	3500001	3566135

Continuation (25 of 36) of AP006840 from base 2400001 (AP006840 Symbiobacterium thermophilus)

Query Match 3.7%; Score 72.6; DB 1; Length 110000;  
Best Local Similarity 42.9%; Pred. No. 6.3;  
Matches 363; Conservative 0; Mismatches 484; Indels 0; Gaps 0;

QY 1068 CCGCGCATGAGGAGCTGACGAGAGAGCTGGCGCCCTTCTCTGCGCAACACGACCA 1127  
D 93382 CTTCTGGGACCGCGCGGCTGGACGCTGGTGTGCTGACCGGTCCTGTATCTCGG 93323  
QY 1128 CAACTTCTGTCCAAAGCGCTCTGTGTACTTCCAGACGCGCTTACTACAGGGGGGCAT 1187  
D 93322 CGGGTGTGTGCTGATGCCGCCCTCGCAGCCCTTACATGGACCAAGTGACCGGCGG 93263

QY 1188 GGGCAAGCTGTGTGCGGTGCTGCAAGTGCCTGGCGGTGCTGTGGACTGGGGCAAGACCGT 1247  
D 93262 GGGCGGGGTGGCCCTGGGTCACTTTCGCGACGCTCGGCTACTGCTGTGCGGAGACGATCGG 93203  
QY 1248 CAAGCGCTCGCCCAACGCGCCCAATGAGAGAGCAGCGCGTCTGTGGGAGCAACAT 1307  
D 93202 CAAGCTGTGGGCAACCGGGAAGACTGCTGTGGAGAAGTACAGTTCCCAAGGGGTCCA 93143  
QY 1308 GCTCATCCACTTCTGTGAAGAACCGGCGCCCAAGCGCTGGTGTGGTGTCTCAAGTTCTGT 1367  
D 93142 GTTCTTCGGGAGTCTGTGATCGCCACCTCGCTGGTGTGATGTTTCATCATGTTCTCATCGC 93083  
QY 1368 GAGCTGTGTCTCTTCAAGCGCGCTGTGTGGTGTGGCGGCGCGCTGTGGGCAAGCA 1427  
D 93082 CTCCTTGGGCGGCGCAACGCGGAGACCTGTGCGGCTCTCGGCGGCGGAGAACATGCTGGT 93023  
QY 1428 GTACGCGCTGATCAAGCGCGGACCGCATCCCAATTGAGAACTACATCGCGCGCAACATGA 1487  
D 93022 CTTCTCGCTCATGACGGCGGTGACCTTTCGCGGCGGTGGCGGTCTGTGCTCTACGGCGT 92963  
QY 1488 CGCGTGGCGGAGAACTCGCACGTGGCGCAAGCAGAACTACTTCTTACTACAACTGCCTCAC 1547  
D 92962 CGCGATGATCATCAACGAGATCGTGGCGGCTTCAAGGGGTTTCGCGGACATCGTGTGCC 92903  
QY 1548 CGGCAAGTTCTTGGCGGCAACCTGCCCCACCTTACCTGCGGAGCGCGGCTTTCGCCACCT 1607  
D 92902 GGACTCAAGCGCGCTGTGACTGCGCGCATCACTTCCCGTACGCGCCACCTGCGGTCT 92843  
QY 1608 CAAGAGTGGGTGGAGCAACCTGACCGTCTCCCACTTCTTCAATGAGAGGCTCAA 1667  
D 92842 GGTGGGCTTCACTGTCAGCTTCTTGGGCGGCTGGTCTCCATGTTTCATCATGAGGCTCTGT 92783  
QY 1668 AGCGCGCACTTACCAAGGTGATTTCTGATGACCACTGCGTGGAGCTGGCTGATATGCCCGT 1727  
D 92782 GGGGCTGCGGTGATCGTGGCGGCTTGTGCGGCTTCTTCTGTTGGCGGCAACGCGCGG 92723  
QY 1728 GGGCAAGAGTGGCGGAGTGTCTGCGGCAAGAGGTTGGCGGCGGCGGATGCTGATCTG 1787  
D 92722 CGTGTTCGGCAACGCGGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 92663  
QY 1788 GCGTCCGCTCCCTCAGCGCGGCTTACGCGGAGCTGATCCAGAGCGGCGGCTTTCGACGT 1847  
D 92662 CATCTCATGAGCTTGGCGGCGGCTTCTGTCGCGGACCTTGGTGGGCTTGGGCTTTCG 92603  
QY 1848 GGGTGCATCCGCGGCGGCACTCAGGGCTTACATGGAGCCGCGTCAACATGTACAGCTCTT 1907  
D 92602 CAACACCACTTGGGCGATGCCGACTTCCAGTGTGGGATCTCTGTCGCGGCGGATCGG 92543  
QY 1908 CTACATG 1914  
D 92542 CAACCTG 92536

Search completed: March 14, 2006, 00:48:18  
Job time : 9934 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 13, 2006, 22:58:40 ; Search time 367 Seconds  
(without alignments)  
9430.279 Million cell updates/sec

Title: US-10-620-914-44

Perfect score: 1947

Sequence: 1 atggggtcgggtcgtagcgg.....gogccaaagaggaactaa 1947

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.\*

- 1: /cgn2\_6/ptodata/1/ina/1 COMB.seq.\*
- 2: /cgn2\_6/ptodata/1/ina/5 COMB.seq.\*
- 3: /cgn2\_6/ptodata/1/ina/6A COMB.seq.\*
- 4: /cgn2\_6/ptodata/1/ina/6B COMB.seq.\*
- 5: /cgn2\_6/ptodata/1/ina/H COMB.seq.\*
- 6: /cgn2\_6/ptodata/1/ina/PTUS COMB.seq.\*
- 7: /cgn2\_6/ptodata/1/ina/PP COMB.seq.\*
- 8: /cgn2\_6/ptodata/1/ina/RE COMB.seq.\*
- 9: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	69.4	3.6	3003	3	US-09-252-991A-14790
2	69.4	3.6	3132	3	US-09-252-991A-14540
3	69.4	3.6	3411	3	US-09-252-991A-14823
4	68.2	3.5	2830	2	US-09-010-928B-1
5	63	3.2	1631	3	US-09-302-540-6564
6	63	3.2	2034	3	US-09-302-540-8514
7	63	3.2	2730	3	US-09-302-540-489
8	63	3.2	6855	3	US-09-302-540-897
9	62.4	3.2	1082	3	US-09-881-165-4
10	62.4	3.2	1167	3	US-09-302-540-2993
11	62.4	3.2	1250	3	US-09-302-540-362
12	62.2	3.2	3579	3	US-09-302-540-7318
13	62.2	3.2	5656	3	US-09-302-540-694
14	61.8	3.2	981	3	US-09-302-540-6237
15	61.8	3.2	13332	3	US-09-302-540-1047
16	61.8	3.2	19726	3	US-09-302-540-1164
17	61.6	3.2	1852	3	US-09-369-852-4
18	61.4	3.2	1281	3	US-09-302-540-7495
19	61.4	3.2	6269	3	US-09-302-540-726
20	60.8	3.1	951	3	US-09-252-991A-10513
21	60.8	3.1	1653	3	US-09-252-991A-10546
22	60.8	3.1	1701	3	US-09-252-991A-10316
23	60.4	3.1	1515	3	US-09-302-540-9184
24	60.4	3.1	9556	3	US-09-302-540-982

25	60.2	3.1	1563	3	US-09-302-540-8878	Sequence 8878, Ap
c 26	60.2	3.1	10528	3	US-09-302-540-945	Sequence 945, App
27	60	3.1	9818	3	US-09-302-540-987	Sequence 987, App
28	59.6	3.1	1941	3	US-09-302-540-6922	Sequence 6922, Ap
29	59.6	3.1	4359	3	US-09-302-540-594	Sequence 594, App
30	59.2	3.0	1773	3	US-09-302-540-7746	Sequence 7746, Ap
31	59.2	3.0	4854	3	US-09-302-540-768	Sequence 768, App
32	58.8	3.0	1293	3	US-09-302-540-6161	Sequence 6161, Ap
c 33	58.8	3.0	1675	3	US-09-302-540-302	Sequence 302, App
34	58.4	3.0	1011	3	US-09-252-991A-12311	Sequence 12311, A
c 35	58.4	3.0	1236	2	US-08-440-856A-1	Sequence 1, Appli
36	57.8	3.0	1526	3	US-09-302-540-5594	Sequence 5594, Ap
c 37	57.8	3.0	41927	3	US-09-302-540-1258	Sequence 1258, Ap
38	57.6	3.0	4345	3	US-09-949-016-2755	Sequence 2755, Ap
39	57.6	3.0	8346	3	US-09-949-016-14497	Sequence 14497, A
40	57.4	2.9	975	3	US-09-365-150-3	Sequence 3, Appli
41	57.4	2.9	2261	3	US-09-158-767-1	Sequence 1, Appli
42	57.4	2.9	2261	3	US-09-713-794-1	Sequence 1, Appli
c 43	57.4	2.9	9191	3	US-09-302-540-918	Sequence 918, App
44	57	2.9	1059	3	US-09-252-991A-7911	Sequence 7911, Ap
45	57	2.9	1338	3	US-09-252-991A-7764	Sequence 7764, Ap

#### ALIGNMENTS

#### RESULT 1

US-09-252-991A-14790  
; Sequence 14790, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 14790  
; LENGTH: 3003  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-14790

Query Match		3.6%	Score 69.4; DB 3; Length 3003;
Best Local Similarity		43.8%	Pred. No. 0.00015;
Matches		454; Conservative	0; Mismatches 571; Indels 12; Gaps 3;
QY	521	TGGGCGTTCGCGACTTCTACGTGAGCGGCAAGTACGACCTGCGCCCTGGCGCAGATGCCT	580
DB	1965	TGGGCGTTCGCGCGAGCAACCCGACCTCAACACCGCGCGCTGCTGATCAACCTCAAGC	2024
QY	581	GGTTCGCGCGCTTCTCTCGCGATCGATCTTCGACATCGACACATTCGATCGCGCCG	640
DB	2025	CGCACAGCGCGCGAGCTCACCGCAGCGAGTGATCCAGCGCTGCAGCGCCGAACTCG	2084
QY	641	AGCGCGCGCGCTTACCTGGAGAGAGAGCTGGAGCGCTGTGGGAGAGAGAACCCAGGGTT	700
DB	2085	ACCACCTGCGCGGATCAAGCTGTATCGACGCGGTGCGGACCTGACCATCGAGACC	2144
QY	701	CGATCCCTACGTGCGTGGTGGCGGCCCTTACTACTGTGTGGATTGGCGGCTGCCA	760
DB	2145	GGGTTCGCGCGACCCAGTACCACTTCCCTTCAGAGCGCCGACCGAGCTGCTGCGCG	2204
QY	761	GGGTTCGCGCGACCCAGTACCACTTCCCTTCAGAGCGCCGACCGAGCTGCTGCGCG	820
DB	2205	AGTGGGTGCGGAGCTGTGGTGGCGGCTGCGAGAGTTGGCGGAGCTGCGGAGCTGCCA	2264
QY	821	TCTGTACGCGAGTCTGTGGAGGAGACCCCGAGCGCGATATGGAGGTGATGAGATCAACC	880

Db 2265 GCAGCTGGCAGGCAAGGGCTTGCAGGCTTACCTGACATCGACGGG-----ACACCG 2318  
Qy 881 CCAAGGACAGGCTGCTGACCTGACCTAGCGGGCTGCAATGCGCTGAACTGCTGTGTC 940  
Db 2319 CTTGGGCTTCCAGCTGCGGCTGAACTCTCCGACATCGACAGCGTGTCTACACGCGCTTCGGCC 2378  
Qy 941 AGGGGGCCGCGAGGTGCTGCGTGGAGTGCACACCGCGCAGTGGCGCTTCTGGAGC 1000  
Db 2379 AGCGCTGATCTGACCATCTTCAACGAGCCACCGAGTACCGGTGTGCTGAGGTGG 2438  
Qy 1001 TGAAGAGGTGGCCATTCAGCAGCTGAGGTTTGAAGACGCTGTGACAGCTGTTCGGCGAGG 1060  
Db 2439 CGCGCAGTTCAGCTCGGCCCGCAGGCCCTTGGAGCAGCTCTAGTCCGCTTCCAGCGACG 2498  
Qy 1061 GCGTGACCCGCGGATGAGAGCTGTACGAGAAAGCTGGCGCCCTTCTGCTGCGAAA 1120  
Db 2499 GCACCCAGGTGCGGCTTGTGAGCTTGGCGAAGGTGGAGAGCGGCATACCTGTGTCGCA 2558  
Qy 1121 CCAGCCACAACTTCTGCTCCAGCGCTCTGCTACTTCCAGCAGCGGCTGTACTACAGG 1180  
Db 2559 TCAACCATATCGCCAGTTTC---CCCTCGGCGACCTGTGTCTTCAACTGGCCAGGGTT 2615  
Qy 1181 GCGCATGGGCAAGCTGTGCTGGGTGCTGCAAGTGCCTGGCGCTGTGCTGCGAATGGGCA 1240  
Db 2616 ACTCCCTGGGCGAGCGGTGAGGCGATCCGTGCGCTCGAGGCCAGCTTGGAGCTGCCG 2675  
Qy 1241 AGACCGTCAAGCGCTCGCCACCGCGCCACATGAGAGGAGCAGCGCGCTGTGAGGACA 1300  
Db 2676 TGAGCATGCGAGGCGAGCTTCGCGCGCGCGCGCTTGGCGCTTGGAGGCTCGCTGTGCA 2735  
Qy 1301 GCAACATGCTCATCC---ACTTCTGTAAGAACGGCGCCAGCGCTGTGCTGCTGTCG 1357  
Db 2736 CGTGTGCTGATCTCGCTCGGTGTGATGACATGATGATGATGATGATGATGATGATG 2795  
Qy 1358 TCAAGTTCTGAGCGCTGTGCTTTCACAAAGGCGCTGTGCTGTGCTGTGCTGTGCTGTG 1417  
Db 2796 AGAGTTTCATCCATCCGCTGACCATCTCTCGACCTCGCTCGCGCGGGGTGGCGCGC 2855  
Qy 1418 CGGCAAGAGTACGCGCTGTATCAAGCGGAGCGGATCCCGATGAGAACTACATCGCGC 1477  
Db 2856 TGCTGGCGCTGATCTGCGCGGCGAGGAGATCGGCATCGTGGCGATCATCGGCATATCC 2915  
Qy 1478 GCACCATGGAGCGGCTGCGGAGAACTCGCAGTGGCGAGCAGAACTACTTCTACTACA 1537  
Db 2916 TGCTGATCGGCATCGTCAAGAAAGCGGATCATGATGATGATGATGATGATGATGATG 2975  
Qy 1538 ACTGCTCACCGCAAG 1554  
Db 2976 AGCGCAAGGAGCAAG 2992

## RESULT 2

US-09-252-991A-14540  
; Sequence 14540, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 14540  
; LENGTH: 3132  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-14540

Query Match 3.6%; Score 69.4; DB 3; Length 3132;  
Best Local Similarity 43.8%; Pred. No. 0.00015;  
Matches 454; Conservative 0; Mismatches 571; Indels 12; Gaps 3;

Qy 521 TGGGCGTTCGCGATCTTCTAGTGAAGCGGCAAGTACGACTGCGCCCTGCGCGCAGATGCGCT 580  
Db 1799 TCGGCGTTCGAGCGGCAACCCGACCTCAACACCGCGCGCTGCTGATCAACCTCAAGC 1858  
Qy 581 GGTGCGCGCTTCTTCTGGCGATCGATCTTCGACATCGACAACTGATGATGATGATGATG 640  
Db 1859 CGCACAGCGAGCGCGAGCTCACCGCAGCGAAGTATCCAGCGCTTGCACCCGAACTCG 1918  
Qy 641 AGCGCGCGCTTACCTTGGAGCAGAACTGAGCGCGTGTGGAGCGAGCAGAAACCCAGGGTT 700  
Db 1919 ACACCTGCGCGGATCAAGCTGTATCATGACGCGGTGACGAGCTTGCACCATCGAGGACC 1978  
Qy 701 CGATCCCTACTGTCGCGTGGCTGCGCGCCCTTACTACGTGTGATGATGATGATGATGATG 760  
Db 1979 GGTGCGCGCACCCAGTACAGTTCACTTGCAGGACCGGACCCGCGAGCTGCTCGCGC 2038  
Qy 761 GCGTTGGCAGCGCTTGCACGAGGAGCGCTGAGCGCGCGCCCATGTTCCCGCCACCT 820  
Db 2039 AGTGGGTGCGCAAGCTGTGCGCGCTGACGAGTTGCGCAGCTGCGCGACGCTCGCCA 2098  
Qy 821 TCTGTACACGAGTCTGTGGAGGACCCCGAGCGGATATGAGAGTGTGATGAGATCAACC 880  
Db 2099 GCGACTGGCAGGACAGGGCTTGCAGGCTTACTTGAACATCGACCGCG-----ACACCG 2152  
Qy 881 CCAAGGACAGGCTGTGACCTGACTAGCGCGCGCTGCAATGCCCTGAACTGCTGTGTCG 940  
Db 2153 CTTGCGCGCTTGGCGTGAAGCTCTCGACATCGACAGCGTGTCTTACACGCTTGGCGC 2212  
Qy 941 AGGGGCGCGCAGGTGCTGCTGGAGTGCACACCCCGCGAGTGGCGCGCTTCTGAGC 1000  
Db 2213 AGCGGTGATCTCGACCATCTTCAACGAGCCACCCAGTACCAGTGTGTGAGGTGG 2272  
Qy 1001 TGAAGAGGTGGCCATTCAGCAGCTGAGTGTGAGGACGCTGTGAGCAGCTGTTCGGCGAGG 1060  
Db 2273 CGCGCGATTCAGCTCGGCCCGCGCCCTTGGAGCAGCTTACTGTCGCTTCCAGCGACG 2332  
Qy 1061 GCGTGACCCGCGCATGAGAGCTGTACGAGAAAGCTGGCGCGCTTCTGTCGCAAA 1120  
Db 2333 GCACCCAGGTGCGCTGTGCGAGCTTGGCGAAGGTGAGAGGAGCGGCATACCTGCTGCGCA 2392  
Qy 1121 CCAGCCACAACTTCTGCTCAAGCGCTCTGCTACTTCCAGCAGCGCTGTACTACAGG 1180  
Db 2393 TCAACCATATCGCCAGTTTC---CCCTCGGCGACCTGTGTCTTCAACTGGCCAGGGTT 2449  
Qy 1181 GCGCATGGGCAAGCTGTGCTGGGTGCTGCAAGTGCCTTGGCGCTGTGCTGAGACTGGGCA 1240  
Db 2450 ACTCCCTGGGAGGCGGTGAGGCGATCCGTGCGCTCGAGGCCAGCTTGGAGCTGCCG 2509  
Qy 1241 AGACCGTCAAGCGCTTCGCCCAACCGGCCCAATGAGAGGAGCAGCGCGCTGTGGGACA 1300  
Db 2510 TGAGCATGCGAGGCGAGCTTCCGCGCGCGCGCTGCGCTTTCGAGGCGCTCGCTGTGCAAC 2569  
Qy 1301 GCAACATGCTCATCC---ACTTCTGTAAGAACGGCGCCCAAGCGCTGTGCTGCTGCTG 1357  
Db 2570 CGTGTGCTGATCTTCGCTCGCTCGGTGTGATGATGATGATGATGATGATGATGATGATG 2629  
Qy 1358 TCAAGTTCTGAGCGCTGTGCTTTCACAAAGGCGCTGTGCTGTGCTGTGCTGTGCTGTGCTG 1417  
Db 2630 AGAGTTTCATCCATCCGCTGACCATCTTCTCGACCTTGCCTCGCGCGGGGTGGCGCGC 2689  
Qy 1418 CGGCAAGAGTACGCGCTGTATCAAGCGGAGCGGCATCCCGATGAGAACTACATCGCGC 1477  
Db 2690 TGCTGGCGCTGATGCTGCGCGGCGAGGAGATCGGCATCGTGGCGATCATCGGCATCATCC 2749  
Qy 1478 GCACCATGGAGCGGTGCGGAGAACTCGCAGCTGGCGAGCAGAACTACTTCTACTACA 1537  
Db 2750 TGCTGATCGGCATCGTCAAGAAAGCGGATCATGATGATGATGATGATGATGATGATGATG 2809



QY 1538 ACTGCTCAGCGCAAG 1554  
Db 2810 AGCGCAACGAGGCAAG 2826

RESULT 3  
US-09-252-991A-14823/c  
; Sequence 14823, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252.991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR FILING DATE: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 14823  
; LENGTH: 3411  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-14823

Query Match 3.6%; Score 69.4; DB 3; Length 3411;  
Best Local Similarity 43.8%; Pred. No. 0.00015;  
Matches 454; Conservative 0; Mismatches 571; Indels 12; Gaps 3;

QY 521 TGGCGGTGCGCACTTCTAGTGGAGCGCAAGTACGACCTGCGCCCTGCGCAGATGCCCT 580  
Db 1349 TCGGCGTGCAGCGCAGCAACCCGACCTCAACACCGCGCCCTGCTGATCAACTCAAGC 1290  
QY 581 GGTGCGCCGTTTCTTCTGCGGATCGATCTTCGACATCGACACATTGACATGCGCCCG 640  
Db 1289 CGCACAGCAGCGCGAGCTCACCGCCAGCGAAGTATCCAGCGCTGACGCCGCAACTCG 1230  
QY 641 AGCGCGCGCTTACCTGAGCAGAGCTGGAGCGCGTGGAGCGGAGCAGACACCCAGGTT 700  
Db 1229 ACACCTGCCGGATCAAGCTGTACATGACGCGGTGAGGACCTGACCATCGAGACC 1170  
QY 701 CGATCCCTACGTGCGGTGCTGCGCCCTTACTACGTGTGATGGCGGCTGCGCA 760  
Db 1169 GGGTGGCGCCGACCACTACGATTCACCTTGCAGGAGCGGACCGGAGCTGCTGCGCG 1110  
QY 761 GCGTGGCCAGCGCTGACAGAGCGGTGAGCGCGGCGCCATGTTCCCGCCCACT 820  
Db 1109 AGTGGGTGCCGAAGCTGTGCGCGGCTGCGAGGATTCGCGCAGCTGCGCGACGTGCCA 1050  
QY 821 TCCTGTACAGCGAGTGTGCGGAGGACCCCGAGCGGATATGGAGGTGATGGAGATCAACC 880  
Db 1049 GCAGCTGGAGGACAGAGGCTTGCAGGCTTACCTGAACTGACCGG-----ACACCG 996  
QY 881 CCAAGGACAGCGTGTGACCTGACTAGCGCGGCTGCAATGCCCTGAACTGCTGCTGTC 940  
Db 995 CTTGCGGCTTGGGTGAACTCTCCGACATGACAGCGGTCTTACAGCGCTTGGCC 936  
QY 941 AGGGGCGCGGAGGTGTGCTGCTGAGTGTGACAGTGTGACAGCGCGGAGTGTGCTGAGC 1000  
Db 935 AGCGGCTGATCTGACCATCTTCAACCGAGGCGGAGTACCGGCTGCTGAGGAGTGG 876  
QY 1001 TGAAGAAGGTGGCATTTACAGCTGAGTGTGAGGAGCGGTGCGAGCTGTTGCGGCGAGG 1060  
Db 875 CGCGGAGTTCAGCTGCGCGGAGGCGCTTGGAGAGCTTACGTGCGGCTCAGGCGAGC 816  
QY 1061 GCGTGCACCGCGCATTTGAGGAGTGTACAGAGAGAGTGGCGGCTTCTGCTGCGCAA 1120  
Db 815 GCACCGAGGTGCGCTGCTGAGCTTGGGAGGTGGAGGAGCGCATACCTCTCTGCGCA 756  
QY 1121 CGAGCCACAACTTCTGCTGCAAGCGGCTCTGGTATCTTCCAGCAGCGGCTGTATCAAGG 1180

Db 755 TCAACCATATCGCCCGATTTC---CCCTCGCGGACCCCTGCTGTTCAACCTGCGCCAAAGGTT 699  
QY 1181 GCGGCATGGGCAAGCTGTGCTGGGTGCTGAGTGCCTGCGCGCTGCTGCTGCGCACTGGGCA 1240  
Db 698 ACTCCCTTGGGCGAGGCGGTTCGAGCGGATCGTGGGCTGAGGCGGAGCTGAGGCTGCGCG 639  
QY 1241 AGACCGTCAAGCGCTGCGCAACGCGCCCAATATGAGGAGGAGCGCGGCTGCTGTTGGGACA 1300  
Db 638 TGAGCATGACAGGCGAGCTTCCGCGGCGCGGCTGCGCTTTCGAGGCGCTGCTGTTGCAACA 579  
QY 1301 GCAACATGCTCATCC---ACTTCTGTAAGAACGGGCCCAAGCCGCTGCTGCTGCTGCTG 1357  
Db 578 CGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 519  
QY 1358 TCAAGTTTCTGAGCTGCTGCTGCTTCAACAAGCGCGCTGCTGCTGCTGCTGCTGCTGCTG 1417  
Db 518 AGAGCTTCACTCCATCGGTCGACCATCTCTCGACCCCTGCGCGGCGGTGCGCGCGC 459  
QY 1418 CGGGCAAGCAGTACGCGCTGATCAAGCGGAGCGGATCCCATTTGAGAACTACATCGCGC 1477  
Db 458 TGCTGCGCTGATGCTGCGCGGCGGAGATCGGATCGTGGCGATCATCGGATCATCC 399  
QY 1478 GCACCATGAGCGGCTGCGGAGAACTCGCACCTGCGCAGCAGCACTACTTCTACTACA 1537  
Db 398 TGCTGATCGGCATCGTCAAGAACGCGGATCATGATGATCGACTTCCGCTCGACCGC 339  
QY 1538 ACTGCTCTCACCGGCAAG 1554  
Db 338 AGCGCAACGAGGCAAG 322

RESULT 4  
US-09-010-928B-1  
; Sequence 1, Application US/09010928B  
; Patent No. 5994099  
; GENERAL INFORMATION:  
; APPLICANT: Lewis, Randolph V  
; APPLICANT: Hayashi, Cheryl Y  
; TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA  
; TITLE OF INVENTION: CODING THEREFOR  
; NUMBER OF SEQUENCES: 29  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH  
; STREET: 8110 GATEHOUSE RD. SUITE 500E  
; CITY: FALLS CHURCH  
; STATE: VIRGINIA  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 22042  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; FILING DATE: 22-JAN-1998  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Murphy Jr., Gerald M  
; REGISTRATION NUMBER: 28977  
; REFERENCE/DOCKET NUMBER: 1447-109P  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2830 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; FEATURE:  
; NAME/KEY:  
; LOCATION: 1..2830  
; OTHER INFORMATION: /note= "Flagelliform DNA sequence"

OTHER INFORMATION: taken from the 5' region. The putative start codon is at  
OTHER INFORMATION: position 219"

FEATURE:  
NAME/KEY: CDS  
LOCATION: 219..2830  
US-09-010-928B-1

Query Match 3.5%; Score 68.2; DB 2; Length 2830;  
Best Local Similarity 46.4%; Pred. No. 0.00025;  
Matches 290; Conservative 0; Mismatches 333; Indels 2; Gaps 2;  
772 GCCTTCACGAGAGCGCTGGAGCGGCCCCCATGTTCCCGCCCACTTCTCTGTACACG 831  
Db |||||  
1033 GACCTGGTGGAGCCGACACAGAGGTTATGACCTGGTGGAGCCGACAGGAGGTTACG 1092  
Qy |||||  
832 CAGTGTGGAGAGACCCGAGCGGATATGAGAGTGATGAGATCAACCCCAAGGACAG 891  
Db |||||  
1093 GACCTGGTGGAGCTGGACACAGGAGTTACGACCTGGTGGAGCTGGGCTTGGAGGTTACG 1152  
Qy |||||  
892 GTCTGACCTGACTAGCGCGGCTGCAATGCCCTGAACCTGCTGGTGACAGGGGCGCGC 951  
Db |||||  
1153 GACCTGGTGGAGCTGGACCTGGAGGTTACGACCTGGTGGAGCTGGACCTGGAGGTTACG 1212  
Qy |||||  
952 CAGTGTGTCTGGTGGACTGCAACCCCGCGCAGTCCGCGCTTCTGGAGCTGAAGAGGTTG 1011  
Db |||||  
1213 GACCTGTGGAACTGGACCTGGTGGATACGACCTGGTGGAACTGGACCTGGAGGTTG 1272  
Qy |||||  
1012 GCATTCAGCAGCTGGAGTTTGGAGAGCTGTGGAGCTGTTCGGCGAGGGCGTGACCCG 1071  
Db |||||  
1273 GACCTGGAGGAGCTGGACACAGGAGATATGACCTGGTGGTGTGGACCTGGTGGTGTG 1332  
Qy |||||  
1072 CGCATTTAGAGCTGTACGAGAGAGAGCTGGCGCCCTTCCTG-TGCAAAACCCAGCCCAA 1130  
Db |||||  
1333 GACCTGTGTGTGGACCTGGTGGTGTGGACCTGGTGGTGTGGACCTGGTGGTGTGCTG 1392  
Qy |||||  
1131 CTTCTGTGTCAAGCGCTCTGTGTACTTCCAGCAGCGGCTGTACTACAGCGCGCATGGG 1190  
Db |||||  
1393 GACCTGTGTGATACGGCCCTTGGTGGATCTGACCTGGTGGTGTGGACCTAGTGGTCCG 1452  
Qy |||||  
1191 CAAGCTGTGTGGTGTGTGAGTCCCTGGCGCTGGTGTGTGGAGCTGGGCAAGCCGTCAA 1250  
Db |||||  
1453 GACTTGTGTGT-GCTGGACCTGGAGGTGCGGAGCTTGTGGAGCAGGACCTGGAGGAGCA 1511  
Qy |||||  
1251 GCGCTCGCCNACGCGCCCAATGAGAGGAGCAGCGCGCTCTGAGGACGACCAATGCT 1310  
Db |||||  
1512 GGAACCAAGTGTGCGGACCCGCTGGAGAGGACCCCGTGGAGCAGGACCAAGTGTGCT 1571  
Qy |||||  
1311 CATCCACTTGTGAAGAACCGGCCCAAGCCGCTGGTGTGGTGTTCGTCAAGTTTCGTGAG 1370  
Db |||||  
1572 GGACCCGCTGGTGGAGGCTGGAGGAGCAGGTGCTGTGTGTGTGTGTGTGTGTGTGCA 1631  
Qy |||||  
1371 CTGGTGTCTTTCAACAGGCGCTG 1395  
Db |||||  
1632 GGTGTGGAGGTGAGGACGTGGTG 1656

## RESULT 5

US-09-902-540-6564  
; Sequence 6564, Application US/09902540  
; Patent No. 6833447  
; GENERAL INFORMATION:  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Wiegand, Roger C.  
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15849)B  
; CURRENT APPLICATION NUMBER: US/09/902,540  
; CURRENT FILING DATE: 2001-07-10  
; PRIOR FILING DATE: 2000-07-10  
; NUMBER OF SEQ ID NOS: 16825  
; SEQ ID NO 6564

LENGTH: 1691  
TYPE: DNA  
ORGANISM: Myxococcus xanthus  
US-09-902-540-6564

Query Match 3.2%; Score 63; DB 3; Length 1691;  
Best Local Similarity 44.9%; Pred. No. 0.0024;  
Matches 323; Conservative 0; Mismatches 390; Indels 6; Gaps 2;

Qy 659 AGCAGAAGCTGGAGCGCGCTGGAGCAGACAGACCCAGGGTTCATCCCTACGTCGGT 718  
Db |||||  
260 AGGTGCGCTGGCGACCGCTGGCGCGAAGCTGCCAACCGCATGATCAAGAGCGCG 319  
Qy |||||  
719 GGTCTGCGCGCCCTACTAGTGTGATTTGGCCGCTGCCAGCGCTTGGCGAGCCCTGC 778  
Db |||||  
320 CGTCTGCGCGCGTGGACAGCTTCGGCATGCTCTGCTCTCTCA---AGAGCTGGGCTGA 376  
Qy |||||  
779 ACAGAGAGCGCTGGAGCGCGCCCATGTTCCCGCCCACTTCTCTGTATACGAGTCGT 838  
Db |||||  
377 GCGAGGAGTCCAGCGGCTGCTCATCTTCCGCGGACACGCGCTGGGACGCGCCATCG 436  
Qy |||||  
839 GCGAGAGCCCGAGCGGATATGAGGTGATGAGATCAACCCCAAGGACACGTCGTA 898  
Db |||||  
437 CGAGAGCGCTGGGCTGGATGGCTGCTGAGGTGAACGTCACGCCGAACCGCCCGG 496  
Qy |||||  
899 CCCTGACTAGCGCGCTGCAATGCCCTGAACCTGTGTGTCAGAGGCGCGCGCAGGTGG 958  
Db |||||  
497 ACCTGCTACGACCTGGGCGTGGCGCGAGGTGAGCGTGGTGGCGGCTCCCGGTTGA 556  
Qy |||||  
959 TGTCTGCTGGAGCTGCAACCCCGCGCAGTTCGCGCTTCTGGAGCTGAAGAGTGGCCATTC 1018  
Db |||||  
557 AGTGTGCGCCAGCGGAGCTCGGGCACGCCGCGCGCGAGCAGGTG---AAGG 613  
Qy |||||  
1019 AGCAGCTGGAGTTTGGAGAGCTGTGCGAGCTGTTCGCGAGGCGCTGCACCCGCGCATTTG 1078  
Db |||||  
614 TGCCTGCTGGAGGCGCGCGACCGCTGCGCGCTACCTGGCGCGCTGGTGGAGAACGTGA 673  
Qy |||||  
1079 AGGAGCTGTACGAGAGAGAGCTGCGCGCTTCTGTGCAAAACAGCCACAACTTCTGGT 1138  
Db |||||  
674 AGATTGCGCGCTCTCGCAGTGTATGAGGACCGCTGTAAGGCGCGCGGCTGGCGCCA 733  
Qy |||||  
1139 CCAAGCGCTCTGTGATTTCCAGCAGCGCTGTACTACAGGCGCGCATGGGCAAGCTGT 1198  
Db |||||  
734 TCAACAAGCTGGTGGATGTCAACAATACCTGAGATACGCGCAGCGCTGCACG 793  
Qy |||||  
1199 GCTGGGTGTGCTGCTGCTGGCTGGCTGGGACCTGGGCAAGACCGTCAAGCGCTCG 1258  
Db |||||  
794 CTTGCACTTGGAGAGCTGGCGGCTCAGGAGATTGCTCGTACGCGCGCGCGCG 853  
Qy |||||  
1259 CCAAGCGCGCCCAATGGAGGAGCAGCGCGCTCTGTGGGACAGCAATGCTCATCCACT 1318  
Db |||||  
854 AGAAGCTCAGAGCTGGAGCGCAAGGACCGCTCTGTGGAGCTGATGCTGTGTCATCG 913  
Qy |||||  
1319 TCCTGAGAGAGCGGCGCCCAAGCCGCTGGTGTGGCTGTTCGTCAGAGCTGGAGCTGGTG 1377  
Db |||||  
914 CGGACAGGACCGCGCGCAGGCGCATCGTGGCGCTCATGCGCGCGCGCGCAGCGAGGTG 972

## RESULT 6

US-09-902-540-8514  
; Sequence 8514, Application US/09902540  
; Patent No. 6833447  
; GENERAL INFORMATION:  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Wiegand, Roger C.  
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15849)B  
; CURRENT APPLICATION NUMBER: US/09/902,540  
; CURRENT FILING DATE: 2001-07-10  
; PRIOR FILING DATE: 2000-07-10  
; NUMBER OF SEQ ID NOS: 16825  
; SEQ ID NO 8514





Db 771 CCAGCGCTCAAGCCCGGAGTACTTGGCAGGCTCGGGCGAGCGGCTCGGCGTGA 830  
Qy 1230 GGAGCTGGGCAAGACCGTCAAGCGCTTCGCAACCGCCCAACAATGAGAGAGAGCGCCG 1289  
Db 831 GGTGCGTGCCTTCCATCTGCGCGCGCAACGCGCGCTCCCAACATGCCCCAGCTCCAGGG 890  
Qy 1290 TCTGTGGGAGCAGCAGTCTATCTCAGTCTGCTGAGAGAGCGGGCCCAAGCGCTGTGTG 1349  
Db 891 GCTGTGGCGCGGTGAGCTGAGTGAAGAGCGGCGCGCTGAGAGAGCTTGGAGCGCTGGAGAC 950  
Qy 1350 GCTGTTCGTCAGTTCGTGAGCGTGTGTCTTCAACAAGCGCGTGTGTGTTCGCGCG 1409  
Db 951 TCGGCGCTCAGTCAGTACTCTTCGCGCGCAAGATCTGACCCGCGCGCGCTCAGGG 1010  
Qy 1410 CGGCGTTCGCGGCAAGCAGTACGCGCTGATCAAGCGCGGAGCGGATCCCAATGAGAACTA 1469  
Db 1011 CGCGTCCAGCAGCATCTCCAACTCGCTGGCGAGCAACCAAGTGAACATCATCCGCGGCT 1070  
Qy 1470 CATCGCGCGCAGCATGAGCGGCTGGCGGAGAACTCGCAGTGGCGGAGAGAACTAC 1527  
Db 1071 GGGACAGGGCGGATGGCGAGGCTTCTCTCGCAAGCAGGTGGCGGTGAAGGGCTTC 1128

## RESULT 11

US-09-902-540-362  
; Sequence 362, Application US/09902540  
; Patent No. 6833447  
; GENERAL INFORMATION:  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Wiegand, Roger C.  
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15849)B  
; CURRENT APPLICATION NUMBER: US/09/902,540  
; CURRENT FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: 60/217,883  
; PRIOR FILING DATE: 2000-07-10  
; NUMBER OF SEQ ID NOS: 16825  
; SEQ ID NO 362  
; LENGTH: 1250  
; TYPE: DNA  
; ORGANISM: Myxococcus xanthus  
US-09-902-540-362

Query Match 3.2%; Score 62.4; DB 3; Length 1250;  
Best Local Similarity 43.7%; Pred. No. 0.003;  
Matches 340; Conservative 0; Mismatches 426; Indels 12; Gaps 1;  
Qy 762 CGTTGGCCAGCCCTGCACGAGGAGCGGTGGAGCGGCCCATGTTCCCGCCACCTT 821  
Db 434 CTTGAGTACGCGGGCGGAGCAGTTCGTGAGTGGACCGGTGCTCATGCGTACCT 493  
Qy 822 CTTGTACCGCATGCTGTGGAGACCCCGGAGATATGAGGTGATGAGATCAACCC 881  
Db 494 GCGCGCGCAGCATGCGCGCGGAGCAGCGGAGTGTGCGACACCTGAAGATCATCCAGGA 553  
Qy 882 CAAGGACAGGCTGTGACCTGTAGTACGCGGCTGCAATCCCTGACTGCTGTGTGCA 941  
Db 554 CGACATCACATATACCTTGGCCAGCGCTGAAAGGGTGAATCAACGTCGCGCGCT 613  
Qy 942 GGGGCGCGGCGAGTGTGTGCTGTGAGTCAACCCCGCGCGAGTCTGCGGCTTCTGGAGCT 1001  
Db 614 GCGTTCGGGCTTGAGGGCGCGGTGCGCTTCGATTCAGCCAGTCAAGAGTGGAGCC 673  
Qy 1002 GAAGAAGGTGCCATTACAGAGCTGGAGTTTGAAGACGTTGGCA-----GCT 1049  
Db 674 GGAGGCCCTCCGAAGCTGAGCAGGTGTGGAGACGGCGGCGCAGGGCGCAGCTGGT 733  
Qy 1050 GTTCGGGAGGGCGTGAACCCGCGCATTTAGAGAGCTGTACGAGAAAGCTTGGCGCCCTT 1109  
Db 734 GTTGTGCGGCGTGCAGCGCGCGGTGTCGCGGTGCTGGCGGCTCGGCCAAGGTGTTGCC 793

Qy 1110 CTTGTGCGAAACACAGCCACAACTTCTGGTTCAGAGCGCTCTGTACTTTCAGAGCGGCT 1169  
Db 794 GGTGCGGCTGGCCACGCTGTGGCTGCGCTGTGACTGCGCAACTGCGGGCAGGTGAGCCA 853  
Qy 1170 GTACTACAGGGCGGATGGCAAGCTGTCTGGTGTCTGCTGCTGCTGCTGCTGCTGCTGCT 1229  
Db 854 CAGCGCTTCAGAGCGCGGATCTCTGGCAGGCTGCGGGCGAGGCTGCGGGCTGGA 913  
Qy 1230 GGGACTGGGCAAGACCGTCAAGCGCTTCGCAACCGGCCCAATGAGAGAGAGCGCGCG 1289  
Db 914 GGTGCGTGCCTCCATCTGCGGGCGCAACCGCGCGTCCCAATGCGCCAGCTCCAGGG 973  
Qy 1290 TCTGTGGGAGCAGCAATGCTCATCTTCTGTGAAGAGAGCGGCCCAAGCGCTGTGTG 1349  
Db 974 GCTGTGCGCGCGGTGAGTCAAGCGCGGCTGAGGAGCTTGGAGCGCTTGGAGAC 1033  
Qy 1350 GCTGTTCGTCAGTTCGTGAGCGTGTCTTCAACAGGCGCTGTGTGTTTCGCGCG 1409  
Db 1034 TCGCGCTCTCAGTCAGTACTCTTTCGCGCGCAGAAATCGACCCGCGCGCTCAGGG 1093  
Qy 1410 CGGCGTTCGCGGCAAGCAGTACGCGCTGATCAAGGGCGGAGCGGATCCCAATTGAGAACTA 1469  
Db 1094 CGGCTCCAGCAGCATCTCCAACTCGCTGGGCGAGCACCAAGTGAACATCATCCGCGCT 1153  
Qy 1470 CATCGCGCGCAGCATGAGCGGCTGGCGGAGAACTCGCAGTGGCGGAGAGAACTAC 1527  
Db 1154 GGGACAGGGCGGATGGCGAGGCTTCTCTCGCAAGCAGGTGGCGGTGAAGGGCTTC 1211

## RESULT 12

US-09-902-540-7318  
; Sequence 7318, Application US/09902540  
; Patent No. 6833447  
; GENERAL INFORMATION:  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Wiegand, Roger C.  
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15849)B  
; CURRENT APPLICATION NUMBER: US/09/902,540  
; CURRENT FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: 60/217,883  
; PRIOR FILING DATE: 2000-07-10  
; NUMBER OF SEQ ID NOS: 16825  
; SEQ ID NO 7318  
; LENGTH: 3579  
; TYPE: DNA  
; ORGANISM: Myxococcus xanthus  
US-09-902-540-7318

Query Match 3.2%; Score 62.2; DB 3; Length 3579;  
Best Local Similarity 42.3%; Pred. No. 0.003;  
Matches 474; Conservative 0; Mismatches 638; Indels 9; Gaps 2;  
Qy 821 TCTGTACACGCGAGTCTGTGGAGACCCCGAGCGCGATATGAGGTGATGAGATCAACC 880  
Db 1517 TCCAGAGCGCAGAGTGTGTGGAGCCCAAGGACGACAAAGGCGAGCGGCGGCGAGC 1576  
Qy 881 CCAAGGACAGGTGTGACCTTCACTAGTACGCGGCTGCAATGCGGCTGAACTCTGTGTGC 940  
Db 1577 CCGGCAAGTGGTGTGAGAACCGCGCTCAACGCGGTGTGGCTTACGACGAGTGTCA 1636  
Qy 941 AGGGGCGCGGAGGTTGTGCTGGTGAATGCAACCCCGCGAGTGGCGCTTCTGGAGC 1000  
Db 1637 AGCGCGCGGAGCGCGGCGGCGGCAAGAGCGAGCGGTGGGCTCGGACATCAACAGA 1696  
Qy 1001 TGAAGAAGTGGCCATTCAGCAGCTGAGTGTGAGGACGTTGTGGAGCTGTTCGGCGAGG 1060  
Db 1697 AGCCACCATCCCGACGCTGAAGAGCGCTGTCTCAAGCGGTGTGAGCGCTACTCAAGT 1756  
Qy 1061 GCGTGACCCGCGCATTTGAGGAGCTGTACGAGAAAGTGGCGGCTTCTCTGTGCAAA 1120

11757 ACSTGCCCAAGGTTGAGAGCGGTGGAGATCGCTTCAAGGGGCGCAACATCTACTACC 1816  
1121 CCAGCCCAACTTCTGTCTCAAGCGCTCTGTACTTCCAGCAGCGGCTGTACTACAGG 1180  
1817 GCCAACACCACTTTCAGCAGCGGCTGTCTGCGCTTACGAGATCGCGCTCGGTACCCCG 1876  
1181 GCGCATGGGCAAGCTGTCTGGTGTGTCAGTGCCTTGGCGGTGTGTCGGAGCTGGCA 1240  
1877 AGTACAGTTTCAGAGCGGCGCGCGGAGATCTCGCCCACTCATCTCGACT 1936  
1241 AGACCGTCAAGCGCTCGCCCAAGCGCCCAATATGGAGAGCAGCGCGTCTGTGGACA 1300  
1937 CGTACCACTCTGTGAGGACTTACGCGAAGGTGAACGAGTGGCGCGCGCTTCTACGCCA 1996  
1301 GCACATGC-----TCATCCACTTCTGTAAGAACGCGGCCCAAGCGCTGTGTGCTGT 1354  
1997 ATGACAGCTGGCGGTGGGCAAGTTCCGCGACGACCTTGGCGAAGCTCATCGAGCGCT 2056  
1355 TCCTCAAGTTTCGTAGCTGTGTCTCTTCAACAGGCGGTGTGTGTTGCGCGGCGCG 1414  
2057 CTTTCAAGCTCTGTGAGGAGTTCGAGGAGAGAGAGGTTTCGAGAGGCGCGCGGCGT 2116  
1415 TGGCGGCAAGCAGTACGCGCTGATCAAGCGGAGCGGATCCCGATTTGAGAACTTACTAT 1474  
2117 ACCTCGCTTCTGTGAGGACTTCCCGCAGACGAGATCGCGACCTTGGCGCTTACAAAG 2176  
1475 GCGCACATGAGAGCGGTGGCGGAGAACTCGACGTGCGCAGCGAGCAGAACTTCTACT 1534  
2177 CGTCTGTCGACTACTCAAGCGGAGCGCTTGGATAGGCGCATCGAGGTGCGCAAGCGCC 2236  
1535 ARACTGCTCCACCGGCAAGTTCTGCGCGCAACTGCGCGCAACTGCGCGAGCGG 1594  
2237 TGTTCGCGGAGTACCGCGCTTCAAGCAGCTGCGCGGACTCCATCTAGCGAAGCGGAGG 2296  
1595 CTTTGGCCACCTCAAGAGTGGGTGGAGCAACTGACCGTCTCCACCAACTTCTTCA 1654  
2297 CGCTGGAGGCACTTGGCGACTTCGAGGACGCGCGGCAAGTACGAGGCGCTTACGTGCGG 2356  
1655 TGGAGGAGCTCAAGCGCGGACCTTACACAGGTGATTCGTATGGACACGTCGATGCGC 1714  
2357 GCTACGAGCGGAGCTTGGGTGACAAAGGCGCAAGCGCGGTGGTGGCGGCAAGAGA 2416  
1715 TGGATATGCGCGTGGCGCAAGCGAGTGGCGGAGTGGCGCAAGCGGTGGCG---GG 1771  
2417 AGCGGGGCGGTGGAGCAAGCAAGCGCGGTGGTGGAGTGGAGTGGAGTGGAGTGGAG 2476  
1772 GCGCATCGTATCGGCGTTCGCGCTTCCCTTACGCGCGCGCTTACGCGAGCTGATCCAGA 1831  
2477 GCGAGTGGCGCTCTTCAAGCGCGCGCACTTACGAGAGGCGCTTGGCGGAGTGAAGCGG 2536  
1832 AGCGGGGCTTGGAGTGGGTGATTCGCGCGCGCACTCAGGGTATCATGGACCGGTCA 1891  
2537 CGTTCGCGCAAGCGGAGCACTTACCTGACGTGTGGCGCGCGGCGAAGGAGCGCGCAAAA 2596  
1892 ACATGTACAGCTCTTCTATATGCGCGCGCGCGGAGGCGCG 1932  
2597 TCCGCTCTCCATCATCGACCTGAGCGGCAAGAGCGGCGCC 2637

## RESULT 13

US-09-902-540-694  
; Sequence 694, Application US/09902540  
; Patent No. 683447  
; GENERAL INFORMATION:  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Wiegand, Roger C.  
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15849)B  
; CURRENT APPLICATION NUMBER: US/09/902,540  
; CURRENT FILING DATE: 2001-07-10

; PRIOR APPLICATION NUMBER: 60/217,883  
; PRIOR FILING DATE: 2000-07-10  
; NUMBER OF SEQ ID NOS: 16825  
; SEQ ID NO 694  
; LENGTH: 5656  
; TYPE: DNA  
; ORGANISM: Myxococcus xanthus  
; US-09-902-540-694

Query Match 3.2%; Score 62.2; DB 3; Length 5656;  
Best Local Similarity 42.3%; Pred. No. 0.0042;  
Matches 474; Conservative 0; Mismatches 638; Indels 9; Gaps 2;

QY 821 TCCTGTACACGCAAGTGTGGAGGACCCCGAGCCGATATGGAGGTGATGAGATCAACC 880  
DB 3573 TCAGAGCGCCAAAGTGTGGAGGCCAAGGACGACAAAGGCAAGCGGCGCAAGC 3632  
QY 881 CCAAGAGACAGGTGTGACCTGACCTAGCAGCGGCGCTGCAATGCCCTGAACTCTGTGTC 940  
DB 3633 CCGGCAAGTGTGTGCAAGCGCGCTTACAAACGCGTGTGGCTTACGACGAGTGTGCA 3692  
QY 941 AGGGGCGCGCCAGGTGTGTGCTGCACTGCAACCCCGCGCAGTGGCGCTTCTGAGC 1000  
DB 3693 AGCGGCGGAGGCGCGGCGGAGGCAAGCGAGGCGGTGGCTTGGACATCACNAGA 3752  
QY 1001 TGAAGAGAGTGGCCATTTCAGCAGCTTGGAGTTTGAAGACGTGTGGCAGCTGTTCGGCGAGG 1060  
DB 3753 AGGCCACCATCCCAACGCTGCTGACGCGGTGTGAGCGCTTACCTCAAGT 3812  
QY 1061 GCGTCAACCGCGCATTTAGGAGCTGTACGAGAGAGTGGCGCGCTTCTGTGCAAA 1120  
DB 3813 ACGTGCCAAGGTTGAGAGCGGTTGAGATCGCTTCAAGCGGCGCAACATCTACTACC 3872  
QY 1121 CAAGCCACAACTTCTGCTCCAGCGCTCTGTACTTCCAGCAGCGCTGTACTACCAGG 1180  
DB 3873 GCCAACCACTTTCAGCAGGCGGTGTGCGCTTTCAGCGAGATCGCGCTCGGTACTCCCG 3932  
QY 1181 GCGCATGTGGCAAGCTGTGTGGTGTGTGAGTGTGCGTGGCGGTGTGTGGGCA 1240  
DB 3933 AGTACAAGTTTCAGAGAGCGGCGCGCGGAGATCTCCGCCAACCTCATCTCGACT 3992  
QY 1241 AGACGCTCAAGCGCTCGCCCAACGCGCCCAATATGAGAGAGCAGCGCGTCTGTGGACA 1300  
DB 3993 CGTACCACTCTGTGAGGACTTACGCGAAGGTGAACGAGTGGCGCGCGCTTCTACGCCA 4052  
QY 1301 GCACATGC-----TCATCCACTTCTGTAAGAACGCGGCCCAAGCGCTGTGTGGCTGT 1354  
DB 4053 ATGACAAGCTGGCGGTGGGCAAGTTTCGCGACGACTTGGCGAAGCTCATCGAGCAGTGT 4112  
QY 1355 TCCTCAAGTTTCGTGAGCTGTGTCTTTCAAAGGCGGTGTGTGTGGTTCGGCGCGCGCG 1414  
DB 4113 CTTTCAAGCTCTGACGCCACTGGAGGAGAAAGAGGTTTCGAGAAAGCGCGCGAGCGT 4172  
QY 1415 TGGCGGCAAGCAGTGTGCTGATCAAGCGGAGCGGCTCCCGATTTGAGAACTTACTATCG 1474  
DB 4173 ACCTGCGCTTCTGTGAAGGACTTCCCGCAGCAGGAGATTCGCGCGCTCTACAAAG 4232  
QY 1475 GCGGCACTATGAGCGGTGGCGGAGAACTCGACGTGCGCAAGCAGAACTACTTCTACT 1534  
DB 4233 CGTCTGTCGACTACTCAAGCGGAGCGCTTGAATAGGCGCATCGAGGTGGCGAAGCGCC 4292  
QY 1535 ACAACTGCTCTCACCGGCAAGTTCTTGTGCGGCAACTTCCGCCACCTACTCTGCGAGCGG 1594  
DB 4293 TGTTCGCGAGTACCGCGCTTCCAGGACGCTGCGGAGTCCATCTACGCGAAGCGGAGG 4352  
QY 1595 CTTTGGCCACCTCAAGAGTGGGTGTGAGAACCTGACGCTTCCACCACTTCTTCA 1654  
DB 4353 CGCTGGAGGCGCAATTTGGCGACTTTCGAGGACGCGCGGCAACGTTACGAGGCGCTTACGTGCGCG 4412  
QY 1655 TGGAGGAGCTCAAGCGCGCACCTTACACCAAGGTGATTTCTGTGAGGACACCTGTCAGTGGC 1714  
DB 4413 GCTACGAGCGGAGCTTGGTGTGCAAGGGCAACGCGAAGGCGGTGTGGCGGCAAGAGA 4472



Qy 1715 TGATATGCGCGGCAACGAGTGGCCGAGTGCCCTGGCCCAAGCAGGTTGCGCC---GG 1771  
Db |||||  
Qy 4473 AGCGGGCGGCTGGAGCAACAAGCGCGGTGGTGACAGAGTGGGACAGTCCAAAG 4532  
Db |||||  
Qy 1772 GCGGCATGTCATCTGGGCTTCGCTCCCTCCTCAGCCGCGCTAGCGGAGCTATCCAGA 1831  
Db |||||  
Qy 4533 CGCAGGTGGCGCTCTCAACCGCGCCACCTACCGAGAGGCGCTGGGGCCAGGTGAAGCGG 4592  
Db |||||  
Qy 1832 AGCGGGCTTTCGACGTGCGTGTGATCGCGCGGCCACTCAAGGCTATCATGACCGGGTCA 1891  
Db |||||  
Qy 4593 CGCTGCGCAACCGGAGCACTACTGACGCTGTGGCGCGCGGAGGAGCGCCGACGAA 4652  
Db |||||  
Qy 1892 ACATGTACAGCTCTTTTCAATGCGCGCGCGCGGAGGCGCC 1932  
Db |||||  
Qy 4653 TCGCCTCTCCATCATCGACTGACGCTGACGCGCAAGAGCGCGCC 4693  
Db |||||

## RESULT 14

US-09-902-540-6237  
; Sequence 6237, Application US/09902540  
; Patent No. 6833447

## GENERAL INFORMATION:

; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Wiegand, Roger C.  
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15849)B  
; CURRENT APPLICATION NUMBER: US/09/902,540  
; CURRENT FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: 60/217,883  
; PRIOR FILING DATE: 2000-07-10  
; NUMBER OF SEQ ID NOS: 16825  
; SEQ ID NO 6237  
; LENGTH: 981  
; TYPE: DNA  
; ORGANISM: Myxococcus xanthus  
US-09-902-540-6237

Query Match 3.2% Score 61.8; DB 3; Length 981;

Best Local Similarity 49.8%; Pred. No. 0.0038;  
Matches 156; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

Qy 1443 GCGGACGGCATCCCCATTGAGAACTACATCGCGCGCACCATGAGCGCGCTGGCGGAGAA 1502  
Db |||||  
Qy 246 GGACGCGAGCGGTACCTCAATCAGCTCATCTGCTGGTGGCAAGCGCGCGCGCGGAGCC 305  
Db |||||  
Qy 1503 CTGCGACGTGGCGGACGAGAACTACTTCTACTAACAATGCTCTCACCGGCAAGTTCTCTGG 1562  
Db |||||  
Qy 306 CCGCGAGACGTTGAAGACGACCGCTGCTGTGTACAGGGCGGCTCCGGGAGCTTCTCTGGC 365  
Db |||||  
Qy 1563 CGAACACTGCCCACTTCTCCACCAACTTCTTCTATGAGGAGCTCAAAAGCGCGCACTACAC 1622  
Db |||||  
Qy 366 GCCACCGCGGACATCCCCCTGGCGACGAGGGCTGGGGCTGGATTTTCGAAGGCGGAGGT 425  
Db |||||  
Qy 1623 GGACAACTGACCGCTTCCACCAACTTCTTCTATGAGGAGCTCAAAAGCGCGCACTACAC 1682  
Db |||||  
Qy 426 CTGCGCATCTCGGCGGACACGCGCCAGGGCAAGAGGCGGAGAACGCGGCGCGGACGCT 485  
Db |||||  
Qy 1683 CAAGGTGATTTCTGATGACCACTGTGATGCTGGCTGGATATGCCCGTGGCCAAAGCTGGC 1742  
Db |||||  
Qy 486 CAAGCTGCTGATGCTGGCCAAATGACGCTGTCCCTGGCGCAACTCATCCCCAACGAGCTGGC 545  
Db |||||  
Qy 1743 CGAGTGCCTGGCC 1755  
Db |||||  
Qy 546 CAAGGGCTTCGAC 558  
Db |||||

## RESULT 15

US-09-902-540-1047  
; Sequence 1047, Application US/09902540  
; Patent No. 6833447

## GENERAL INFORMATION:

; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Wiegand, Roger C.  
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15849)B  
; CURRENT APPLICATION NUMBER: US/09/902,540  
; CURRENT FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: 60/217,883  
; PRIOR FILING DATE: 2000-07-10  
; NUMBER OF SEQ ID NOS: 16825  
; SEQ ID NO 1047  
; LENGTH: 1332  
; TYPE: DNA  
; ORGANISM: Myxococcus xanthus  
US-09-902-540-1047

Query Match 3.2% Score 61.8; DB 3; Length 1332;

Best Local Similarity 49.8%; Pred. No. 0.0058;  
Matches 156; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

Qy 1443 GCGGACGGCATCCCCATTGAGAACTACATCGCGCGCACCATGAGCGCGCTGGCGGAGAA 1502  
Db |||||  
Qy 11288 GGACGCGAGCGGTACCTCAATCAGCTCATCTGCTGGTGGCAAGCGCGCGCGGAGCC 11347  
Db |||||  
Qy 1503 CTGCGACGTGGCGGACGAGAACTACTTCTACTAACAATGCTCTCACCGGCAAGTTCTCTGG 1562  
Db |||||  
Qy 11348 CCGGAGACGTTGAAGACGACGCCGCTGGTGTACAGGGCGGCTCCGGGAGCTTCTCTGGC 11407  
Db |||||  
Qy 1563 CGAACACTGCCCACTTACTGCGGAGGCGGCTTCCGACACCTCAAGAGTGGCGTGGT 1622  
Db |||||  
Qy 11408 GCCACCGCGGACATCCCCCTGGCGGACGAGGCTGGGCTGGATTTTCGAAGGCGAGGT 11467  
Db |||||  
Qy 1623 GGACAACTGACCGCTCTCCACCAACTTCTTCTATGAGGAGCTCAAAAGCGCGCACTACAC 1682  
Db |||||  
Qy 11468 CTGCGCATCTGGGCGACACGCCCCAGGGCACGAAGCGGAGAACGCGGCGCGGACGT 11527  
Db |||||  
Qy 1683 CAAGGTGATTTCTGATGACCACTGAGCTGGCTGGATATGCCCGTGGCCAAAGCTGGC 1742  
Db |||||  
Qy 11528 CAAGCTGCTGATGCTGGCCAAATGACGCTGTCCCTGGCGCAACTCATCCCCAACGAGCTGGC 11587  
Db |||||  
Qy 1743 CGAGTGCCTGGCC 1755  
Db |||||  
Qy 11588 CAAGGGCTTCGAC 11600  
Db |||||

Search completed: March 14, 2006, 01:18:31

Job time : 371 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 14, 2006, 00:48:26 ; Search time 1628 Seconds  
(without alignments)  
9889.729 Million cell updates/sec

Title: US-10-620-914-44

Perfect score: 1947  
Sequence: 1 atggggctcggtcggtgacgg.....gcccgaagaggacaactaa 1947

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_NA\_Main:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US11\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1947	100.0	1947	7	US-10-620-914-44
2	645	33.1	5242	7	US-10-620-914-43
3	100.2	5.1	1252	5	US-10-118-495-1
4	100.2	5.1	1252	7	US-10-620-914-1
5	76.8	3.9	2259	8	US-10-741-849-6021
6	73	3.7	2727	7	US-10-620-914-49
7	73	3.7	3427	7	US-10-620-914-48
8	72.4	3.7	137560	8	US-10-481-112-1
9	70	3.6	1251	5	US-10-118-495-32
10	70	3.6	1251	7	US-10-620-914-32
11	69.6	3.6	1248	5	US-10-118-495-22
12	69.6	3.6	1248	7	US-10-620-914-22
13	67.8	3.5	1104	6	US-10-204-434A-12
14	67.8	3.5	6798	3	US-09-918-740-57
15	67.8	3.5	6798	6	US-10-204-434A-13
16	67.8	3.5	6798	8	US-10-835-516-57
17	67.8	3.5	6798	10	US-11-053-541-57
18	67.8	3.5	8077	3	US-09-918-740-63
19	67.8	3.5	8077	8	US-10-835-516-63
20	67.8	3.5	8077	10	US-11-053-541-63
21	67.6	3.5	5858	8	US-10-488-056-14
22	66.6	3.4	1626	7	US-10-392-041-3
23	66.6	3.4	1626	7	US-10-394-763-3

ALIGNMENTS

RESULT 1  
US-10-620-914-44  
; Sequence 44, Application US/10620914  
; Publication No. US20040093639A1  
; GENERAL INFORMATION:  
; APPLICANT: Benning, Christoph  
; APPLICANT: Riekhof, Wayne  
; APPLICANT: Klug Rouven  
; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids  
; FILE REFERENCE: MSU-07769  
; CURRENT APPLICATION NUMBER: US/10/620,914  
; CURRENT FILING DATE: 2003-07-16  
; PRIOR APPLICATION NUMBER: 10/118,495  
; PRIOR FILING DATE: 2002-04-08  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: Patent in version 3.2  
; SEQ ID NO 44  
; LENGTH: 1947  
; TYPE: DNA  
; ORGANISM: Chlamydomonas reinhardtii  
US-10-620-914-44

Query Match	Best Local Similarity	Score 1947;	DB 7;	Length 1947;
Matches 1947;	Conservative	0;	Mismatches	0; Indels 0; Gaps 0;
Qy	1	ATGGGGTCTGGGTCTGTCAGCGCGCGCTGGAGCTACACCAAGAGAACTTCTCCCTGGAG	60	
Db	1	ATGGGGTCTGGGTCTGTCAGCGCGCGCTGGAGCTACACCAAGAGAACTTCTCCCTGGAG	60	
Qy	61	AAGCTCAAGCTCAGCAGCATGAAGGATGACCTGACCGCTTCTGCGCCATATGTGTTCCGC	120	
Db	61	AAGCTCAAGCTCAGCAGCATGAAGGATGACCTGACCGCTTCTGCGCCATATGTGTTCCGC	120	
Qy	121	AGCAAGAGGGCGGATGATCAGCTGCTCGCTGGAGAGCTTCTACGGGCCCCCAGGCGCT	180	
Db	121	AGCAAGAGGGCGGATGATCAGCTGCTCGCTGGAGAGCTTCTACGGGCCCCCAGGCGCT	180	
Qy	181	GCCTTTGCTGCCCGCTGCGCGAGCGCTCGAACCTCATCTGGGTGACCTGGGTGGTGGC	240	
Db	181	GCCTTTGCTGCCCGCTGCGCGAGCGCTCGAACCTCATCTGGGTGACCTGGGTGGTGGC	240	
Qy	241	ACTGGGGAGATGTCGATATGATGGCTGATTACATGACCTGCGGAAGTTCAAGTCCATC	300	
Db	241	ACTGGGGAGATGTCGATATGATGGCTGATTACATGACCTGCGGAAGTTCAAGTCCATC	300	



QY 1165 GGCCTGTACTACAGGGCGGCATGGCAAGCTGTGCTGGGTGCTGCAAGTCCCTGGCGGTG 1224  
 DB 3348 GGCCTGTACTACAGGGCGGCATGGCAAGCTGTGCTGGGTGCTGCAAGTCCCTGGCGGTG 3407  
 QY 1225 GTCTGGGACTGGGCAAGAGCGCTCAGCGGCTCGCCAAACCGGCCCAATGGAGGAGCAG 1284  
 DB 3408 GTCTGGGACTGGGCAAGAGCGCTCAGCGGCTCGCCAAACCGGCCCAATGGAGGAGCAG 3467  
 QY 1285 CGCGCTGTGGGACAGCAATGCTCATCACTTCTGTAAGAAACGGGGCCCAAGCGCGCTG 1344  
 DB 3468 CGCGCTGTGGGACAGCAATGCTCATCACTTCTGTAAGAAACGGGGCCCAAGCGCGCTG 3527  
 QY 1345 GTGTGGCTGTGTCAAGTTCTGAGGCTGTGCTCTTCAACAAAGCGCGTGTGTGTTTC 1404  
 DB 3528 GTGTGGCTGTGTCAAGTTCTGAGGCTGTGCTCTTCAACAAAGCGCGTGTGTGTTTC 3587  
 QY 1405 GGGGGGGCGGTGGGGGAGGAGTACCGGCTCATCAAGGGGAGCGGCATCCCATTTAG 1464  
 DB 3588 GGGGGGGCGGTGGGGGAGGAGTACCGGCTCATCAAGGGGAGCGGCATCCCATTTAG 3647  
 QY 1465 AACTATACATCGCGGCGCATATGAGCGGCTGGCGGAGAACTCGCACTGGCGCAAGCAGAAC 1524  
 DB 3648 AACTATACATCGCGGCGCATATGAGCGGCTGGCGGAGAACTCGCACTGGCGCAAGCAGAAC 3707  
 QY 1525 TACTTCTACTACAACTGCTTCAACGGCAAGTTCTTGGCGGACAACTGCCCCACCTACCTG 1584  
 DB 3708 TACTTCTACTACAACTGCTTCAACGGCAAGTTCTTGGCGGACAACTGCCCCACCTACCTG 3767  
 QY 1585 CGCGAGGCGGCTTGGCCACCTCAGAGTGGGTGTGAGCAACCTGACCGTCTCCACC 1644  
 DB 3768 CGCGAGGCGGCTTGGCCACCTCAGAGTGGGTGTGAGCAACCTGACCGTCTCCACC 3827  
 QY 1645 AACTTCTTATGAGGAGCTCAAGCGCGCACTTACACCAAGGTG 1689  
 DB 3828 AACTTCTTATGAGGAGCTCAAGCGCGCACTTACACCAAGGTG 3872

## RESULT 3

US-10-118-495-1  
 ; Sequence 1, Application US/10118495  
 ; Publication No. US20030074688A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Benning, Christoph  
 ; APPLICANT: Riekhof, Wayne  
 ; APPLICANT: Klug, Rouven  
 ; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids  
 ; FILE REFERENCE: MSU-06897  
 ; CURRENT APPLICATION NUMBER: US/10/118,495  
 ; CURRENT FILING DATE: 2002-04-08  
 ; PRIOR APPLICATION NUMBER: 60/283,812  
 ; PRIOR FILING DATE: 2001-04-13  
 ; NUMBER OF SEQ ID NOS: 42  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 1  
 ; LENGTH: 1252  
 ; TYPE: DNA  
 ; ORGANISM: Rhodobacter sphaeroides  
 US-10-118-495-1

Query Match 5.1%; Score 100.2; DB 5; Length 1252;  
 Best Local Similarity 50.9%; Pred. No. 2e-16;  
 Matches 299; Conservative 0; Mismatches 273; Indels 15; Gaps 2;  
 QY 809 TCCCGCCACCTTCTGTACACGAGTCTGTGGAGGACCCGAGCGGATATGAGGTGA 868  
 DB 122 TCTTCCACGGCCTCGTCTATCCGAGATCTGGGAGGATCCGGCGGTGGACATGGCGGCC 181  
 QY 869 TGGAGATCAACCCCAAGGACACGGTGTGCTGACCTGTAGCGCGGTGCAATGCCCTGA 928  
 DB 182 TCGCCATCCGCCCCGGGACCGGCTGTGGCCATCGCTCGGGCGGTGCAAGTGTCTT 241  
 QY 929 ACCTGTGTGAGGGGGCGGCAAGTGTGTGCTGAGTGTGCTGCAACCCCGCGAGTCGG 988

DB 242 CCTATCTCAGCAGGGGCGGGCTCGATCTCGCGTGGATCTCTCGCCCGCCCATGTGG 301  
 QY 989 CGCTTCTTGAGAGTGAAGAGTGGCCATTTCAGCAGCTG---GAGTTTGAAGAGCTGTGGC 1045  
 DB 302 CGCTGGGGCGGCTGAAGCTCGCCCGCGGGAGAGCTTCCCGACCATGCGCCCTTCTTCG 361  
 QY 1046 AGCTGTTCGGCGAGGCGGTGCACCCCGCGCATTTGAGAGTGTACGAGAAAGAGTGGCGC 1105  
 DB 362 ATCTCTTCGGTTCGCGCAGAGCTCGCCCGGCAATCGCGCCCTCTACGACCGGCACATCGCGC 421  
 QY 1106 CTTCTGTGTGCGAAACAGCAGCAAACTTCTGTGTGTCAGAGCGCCTCTGGTACTTCCAGCAG 1165  
 DB 422 CCGCGCTCGACGGCCGGAGCGCGCTACTTCTGAGGAGCGCGCAGCCCTTCCGGCGGGCGCA 481  
 QY 1166 GCCTGTACTACAGAGGCGCGCATGGGCAAGCTGTGTGGGTGCTGCAAGTGTCTGCGCGG--- 1222  
 DB 482 TCCAGCTGTTCGAGCGGGCTTCTACCGGACGGTGGCTCGGCGCGCTTCTATCGGCGGG 541  
 QY 1223 -----TGTGTGGGACTGGGCAAGACCGTCAAGCGCCTCGGCAACGCCGCCCA 1273  
 DB 542 CCATACGCTCGCGCGGGCGGGCACGACCTGCGGGGCTTCTCGACTGTCCCGACA 601  
 QY 1274 TGGAGAGCAGCGCGCTCTGTGGGACAGCAACATGTCTATCCACTTCTGTGAGAACGGGC 1333  
 DB 602 TCGAGGCGCAGCGCAGCTTCTTCTACGCCCATATCGGGCGGCTCTTTCGAGGCGCGCGTGG 661  
 QY 1334 CCAAGCGGCTGTGTGGCTTTCGTCAAGTTCTGTGAGCTGTGTGCTC 1380  
 DB 662 TGCAGGCGCTTCGCCCGACCGCGCGCGCTCTTTCGGGTGGGGATC 708

## RESULT 4

US-10-620-914-1  
 ; Sequence 1, Application US/10620914  
 ; Publication No. US20040093639A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Benning, Christoph  
 ; APPLICANT: Riekhof, Wayne  
 ; APPLICANT: Klug, Rouven  
 ; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids  
 ; FILE REFERENCE: MSU-07769  
 ; CURRENT APPLICATION NUMBER: US/10/620,914  
 ; CURRENT FILING DATE: 2003-07-16  
 ; PRIOR APPLICATION NUMBER: 10/118,495  
 ; PRIOR FILING DATE: 2002-04-08  
 ; NUMBER OF SEQ ID NOS: 52  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 1  
 ; LENGTH: 1252  
 ; TYPE: DNA  
 ; ORGANISM: Rhodobacter sphaeroides  
 US-10-620-914-1

Query Match 5.1%; Score 100.2; DB 7; Length 1252;  
 Best Local Similarity 50.9%; Pred. No. 2e-16;  
 Matches 299; Conservative 0; Mismatches 273; Indels 15; Gaps 2;  
 QY 809 TCCCGCCACCTTCTGTACACGAGTCTGTGGAGGACCCGAGCGGATATGAGGTGA 868  
 DB 122 TCTTCCACGGCCTCGTCTATCCGAGATCTGGGAGGATCCGGCGGTGGACATGGCGGCC 181  
 QY 869 TGGAGATCAACCCCAAGGACACGGTGTGCTGACCTGTAGCGCGGTGCAATGCCCTGA 928  
 DB 182 TCGCCATCCGCCCCGGGACCGGCTGTGGCCATCGCTCGGGCGGTGCAAGTGTCTT 241  
 QY 929 ACCTGTGTGAGGGGGCGGCAAGTGTGTGCTGAGTGTGCTGCAACCCCGCGAGTCGG 988  
 DB 242 CCTATCTCAGCAGGGGCGGGCTCGATCTCGCGTGGATCTCTCGCCCGCCCATGTGG 301  
 QY 989 CGCTTCTGAGGCTGAAGAGTGGCCATTCAGCAGCTG---GAGTTTGAAGAGTGTGGC 1045  
 DB 302 CGCTGGGCGGCTGAAGCTCGCGCGCGGACGCTGCGCCGACCATGCGCGCTTCTTCG 361





```
; SEQ ID NO 48
; LENGTH: 3427
; TYPE: DNA
; ORGANISM: Neurospora crassa
US-10-620-914-48

Query Match      3.7%; Score 73; DB 7; Length 3427;
Best Local Similarity 51.0%; Pred. No. 4.8e-09;
Matches 172; Conservative 0; Mismatches 165; Indels 0; Gaps 0;

QY 779 ACCGAGAGCGCGTGGAGCGCGCCCATGTTCCCGCCACCTTCTCTGTACACGCGTGTG 838
DB 1906 ACCGAGATCAACTCCCGAAGCACCACCGATTCATGACGAGTACATCTACGCTTTACCT 1965

QY 839 GCGAGACCCCGAGCCCGATATGAGGTGATGAGATCAACCCCAAGGACACGCGTGTGA 898
DB 1966 GCGAAGACTCGCGGTGCGACAGAGAACTCTTAACCTCGGGCCGCGACGCGTCTCTAG 2025

QY 899 CCTGACTAGCGGGGTGCAATGCCCTGAACTGCTGTGCGAGGGGCGCGCAGGTGG 958
DB 2026 CCATCACCAGCGCGCGCAACATTTCTTCTACCTGATGCGAGGTCCGCTCGCGTGC 2085

QY 959 TGTCTGCTGGAAGTCAACCCCGCGCAGTCCGCGCTTCTGAGCTGAAGAAGTGGCCATTC 1018
DB 2086 ACCGCAATCGACCTAAACCCAGCCCAAAACCACTGCTTGAACCAAGTCCGCTCTTTTA 2145

QY 1019 AGCAGCTGGAAGTTGAGGAGCTGTGGCAGCTGTTCGCGGAGGCGGTGCACCCGCGCATTG 1078
DB 2146 CGACTCTGGATTACCCCGAGCTGTGGAAGATCTTCGTTGAGGCGAAACACCCCGACTTTC 2205

QY 1079 AGGAGCTGTACGAGAAGAGCTGGCGCCCTTCTCTGTC 1115
DB 2206 GCTCACTGCTCATCTCCAAACTCTCCCTCACCTCTC 2242

RESULT 8
US-10-481-112-1
; Sequence 1, Application US/10481112
; Publication No. US20040235721A1
; GENERAL INFORMATION:
; APPLICANT: WEBER, Olaf
; APPLICANT: FRIEDERICH, Sonja, Maria
; APPLICANT: SIEGLING, Angela
; APPLICANT: SCHLAPP, Tobias
; APPLICANT: MERCER, Andrew, Allan
; APPLICANT: FLEMING, Stephen, Bruce
; TITLE OF INVENTION: Recombinant proteins of Parapoxvirus ovis and pharmaceutical comp
; FILE REFERENCE: Lea 35228
; CURRENT APPLICATION NUMBER: US/10/481,112
; CURRENT FILING DATE: 2003-12-12
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 137560
; TYPE: DNA
; ORGANISM: Parapoxvirus ovis NZ2
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)...(539)
; OTHER INFORMATION: ORF: L1
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (781)...(449)
; OTHER INFORMATION: ORF: L2r
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1933)...(1664)
; OTHER INFORMATION: ORF: L3r
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3269)...(2790)
; OTHER INFORMATION: ORF: L4r
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2799)...(3851)
; OTHER INFORMATION: ORF: L5
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2962)...(3753)
; OTHER INFORMATION: ORF: L6
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3784)...(3122)
; OTHER INFORMATION: ORF: L7r
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (4341)...(4129)
; OTHER INFORMATION: ORF: L8r
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (4904)...(4428)
; OTHER INFORMATION: ORF: L9r
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (6517)...(4970)
; OTHER INFORMATION: ORF: L1r
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (8042)...(6684)
; OTHER INFORMATION: ORF: L2r
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (9989)...(8070)
; OTHER INFORMATION: ORF: L3r
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (11195)...(10062)
; OTHER INFORMATION: ORF: L4r
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (11493)...(11227)
; OTHER INFORMATION: ORF: L5r
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (11802)...(12038)
; OTHER INFORMATION: ORF: L6
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (12358)...(12080)
; OTHER INFORMATION: ORF: L7r
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (13980)...(12364)
; OTHER INFORMATION: ORF: L8r
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (14826)...(14053)
; OTHER INFORMATION: ORF: L9r
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (15080)...(15394)
; OTHER INFORMATION: ORF: L10
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (16838)...(15423)
; OTHER INFORMATION: ORF: L11r
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (19021)...(16847)
; OTHER INFORMATION: ORF: L12r
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (19704)...(19156)
; OTHER INFORMATION: ORF: L13r
; FEATURE:
```

```
/ NAME/KEY: CDS
/ LOCATION: (20314)...(19736)
/ OTHER INFORMATION: ORF: 14r
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (20401)...(22101)
/ OTHER INFORMATION: ORF: 15
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (22125)...(22940)
/ OTHER INFORMATION: ORF: 16
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (23003)...(23866)
/ OTHER INFORMATION: ORF: 17
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (26908)...(23873)
/ OTHER INFORMATION: ORF: 18r
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (26926)...(27213)
/ OTHER INFORMATION: ORF: 19
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (27626)...(27216)
/ OTHER INFORMATION: ORF: 20r
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (29754)...(27616)
/ OTHER INFORMATION: ORF: 21r
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (32217)...(29800)
/ OTHER INFORMATION: ORF: 22r
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (33380)...(32418)
/ OTHER INFORMATION: ORF: 23r
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (33602)...(33393)
/ OTHER INFORMATION: ORF: 24r
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (34466)...(33612)
/ OTHER INFORMATION: ORF: 25r
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (34735)...(34502)
/ OTHER INFORMATION: ORF: 26r
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (35905)...(34739)
/ OTHER INFORMATION: ORF: 27r
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (37194)...(35905)
/ OTHER INFORMATION: ORF: 28r
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (37200)...(39248)
/ OTHER INFORMATION: ORF: 29
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (41037)...(39229)
/ OTHER INFORMATION: ORF: 30r
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (41374)...(42066)
/ OTHER INFORMATION: ORF: 31
/ FEATURE:
/ NAME/KEY: CDS
```

---

```
/ LOCATION: (42336)...(41731)
/ OTHER INFORMATION: ORF: 32r
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (42407)...(41997)
/ OTHER INFORMATION: ORF: 33r
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (42410)...(43765)
/ OTHER INFORMATION: ORF: 34
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (43770)...(43958)
/ OTHER INFORMATION: ORF: 35
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (43980)...(44534)
/ OTHER INFORMATION: ORF: 36

Query Match      3.7%; Score 72.4; DB 8; Length 137560;
Best Local Similarity 42.9%; Pred. No. 1.5e-08;
Matches 461; Conservative 0; Mismatches 611; Indels 2; Gaps 2;

Qy 618 CGACAAATTGACATCGGCCCCGAGCGCGCCTTACCTGAGAGCAGAGCTGGAGCGCGT 677
Db 20613 CGAGGTCTTCGACGCGGTGCTCAAGCGCGCTTCGACGCGCCTCAGCTCTTTGTGCGGGC 20672
Qy 678 GTGGGAGCAGAACACACCGGTTTCGATCCCTACGTGCGGTGGCTGCGGCCCTACTA 737
Db 20673 GCTCTCGGCGCACTTCGAGGCGCTTCGTTGCCGCATCCGCGGCTACCGCGCGACCTCGC 20732
Qy 738 CGTGTGATTTGGCGCGCTGCCAGCGTTGGCCACGCGCTGCACGAGAGAGCGGTGGAGCG 797
Db 20733 GGAGCTGCTGGCCGACGCGCGCTTCATGAGTGGCTGCGCGCGCGCGAGCTCGCGGA 20792
Qy 798 GCCGCCCATGTTCCCGCCACCTTCTGTGTACGCGAGTCTGTGGAGAGACCCGAGCCGGA 857
Db 20793 GGTCAATTGGCGTGAACACGACATCGCGCGAACCGCTCTTCGCGGACGCGGCGCGCGT 20852
Qy 858 TATGGAGGTGATGAGATCAACCCCAAGGACACGCTGCTACCTGACTAGCGCGCGCTG 917
Db 20853 GCGGACGCGGAGCTCATTTTCGCAAGACCTTCGCGAAGACGAGTTCGCGCGCGTCAA 20912
Qy 918 CAATGCCCTGAACCTGCTGTGTCAGAGGGGCGCGCACAGTGGTGTGCTGAGTGCACACCC 977
Db 20913 GCGCTCGCGCTGCTGCGGTGCTGTGCTGGGCTTCTCTGTGAAGAAGACCTTGGCGG 20972
Qy 978 CGCGCAGTTCGCGCGCTTCTGAGAGTGAAGAAGGTGGCCATTCAGCAGCTGGAGTTGAGGA 1037
Db 20973 CGAGTACGCGGACAAACGACCGCCAGGACCTGTTTACGCTGCTGCAGAAGGCGCGGGGCC 21032
Qy 1038 CGTGTGCGAGCTGTTTCGCGAGGGCGGTGCACCGCGCATTTGAGGAGCTGTACGAGAAGAA 1097
Db 21033 CGTGCGCCA-CAGCGCGCTCACAGAGAGCATCCGCGAGTACCTTCTCCCGGAGACAGGC 21091
Qy 1098 GCTGGCGCGCTTCTGTCGCAAAACACGACCACTTCTGTGTCACAGCGCCT-CTGGTACT 1156
Db 21092 CCAGCCACTGGGTCTGGCTGAAACGCGCGCTGGCGGACGCGAGAGGTGTACCGCGACC 21151
Qy 1157 TCCAGCACGCGCTGTACTACAGGGCGGCAATGGGCAAGCTGTGCTGGGTGCTGAGTGGCC 1216
Db 21152 GGGCGGCGCGCAGCTCTACGAGCGCGTGTCTCAGCTACGCGGTATCAGAGGTCAAGCAGG 21211
Qy 1217 TGGCGGTGGTGTGGGACTGGGCAAGACCGTCAAGCGCTTCGCCAAACGCGCCCAATGG 1276
Db 21212 GGGCGGTGAACGCGCAACACGCTCAAGCTCGTGTACCGGCTCGAGAGACACCCGACATCA 21271
Qy 1277 AGGAGCAGCGCGCTCTGTGGGACAGCAACATGCTCATCTCCATCTCGTGAAGAACGGGCCA 1336
Db 21272 AGGTCTGTGCTGCAGCTCATCTACGCGTTCGCGCGGACATGTCGGGCTCGTGAGT 21331
Qy 1337 AGCGCTGGTGTGCTGTTTCGTTCAAGTTCGTGAGCGCTGGTGTCTTTCACAGGCGCGTGC 1396
```

Db 21332 CCGGAAACGAGAGTGGCGGAGCTACTCTCTGTAGTCTGTACCGAGAACTTCGTGACG 21391  
QY 1397 TGTGGTTCCGCGCGCGGTCCCGGCAAGCAGTACGCGCTGTATCAAGCGGAGCGGATCC 1456  
Db 21392 GACGACCTTCACTCGGACGCGGCTTCGCGAGACCTTTCGCGTGTGCGCGCG 21451  
QY 1457 CATTGAGAACTACATCGCGGACCACTGAGCGGCTGGCGGAGAACTCGCACGTGGCA 1516  
Db 21452 TCGATCCGCACTTCTTCGAGCCGAGGCGCATCGCGAGGCTTCAGCGCAGACGCGCGC 21511  
QY 1517 AGCAGAACTACTTCTACTCAACTGCTCTACCGCAAGTTCCTGCGGACAACTGCCCCA 1576  
Db 21512 TCGAGAGCGCTTACGCGATGAGCTCAACAGCGCTTATGTGCGACCTCATTCAG 21571  
QY 1577 CTTACTGCGCGAGGCGCTTCGCCACCCCTCAAGAGTGGCGTGGTGAACAACCTGACCG 1636  
Db 21572 ACTCGGTGGACCCGCGCTCGCGCGCGCGGCTCGCACTGCGGTGCACACG 21631  
QY 1637 TCTCCAGCAACTTCTTCATGAGAGAGCTCAAGCGCGCACTTACCAAGGTGA 1690  
Db 21632 AGGACTCCGACTACTTTCATCCGGGAGTACCAACCTACTCTTCTCCTCAGCGAGA 21685

## RESULT 9

US-10-118-495-32  
Query Match 3.6%; Score 70; DB 5; Length 1251;  
Best Local Similarity 52.8%; Pred. No. 2.5e-08;  
Matches 151; Conservative 0; Mismatches 135; Indels 0; Gaps 0;  
Publication No. US20030074688A1  
GENERAL INFORMATION:  
APPLICANT: Benning, Christoph  
APPLICANT: Riekhof, Wayne  
APPLICANT: Klug, Rouven  
TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids  
FILE REFERENCE: MSU-06897  
CURRENT APPLICATION NUMBER: US/10/118,495  
CURRENT FILING DATE: 2002-04-08  
PRIOR APPLICATION NUMBER: 60/283,812  
PRIOR FILING DATE: 2001-04-13  
NUMBER OF SEQ ID NOS: 42  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 32  
TYPE: DNA  
ORGANISM: Sinorhizobium meliloti  
US-10-118-495-32

QY 806 TGTTCGCGCCACCTTCTGTACACGAGTCTGTGGGAGGACCCCGAGCGGATATGGAGG 865  
Db 116 TGCTCTTTCCGAGCTCGTCTACCGCAGATCTGGGAGGACCGATTGTGACATGGAAG 175  
QY 866 TGATGAGATCAACCCCAAGACACGGTGTGACCTGTAGTACGCGGCTGCAATGCC 925  
Db 176 CGATGAGATCCGCTCCGAGACATCGGATCTGACGATCGGTTCCGCGGCTGCAACATGC 235  
QY 926 TGAACCTGTGTTGAGGCGGCGCGGAGTGTGTGCTGAGTGTGAGGAGTGTGGC 985  
Db 236 TGACCTATCTCTCCGCGAGCTTCGCCGATAGACGTGTGATCTCAACCCCATCA 295  
QY 986 CGCGCTTCTGAGGCTGAAGAGTGGCCATTCAGCAGCTGGAGTTTGGAGAGTGTGGC 1045  
Db 296 TCGCGCTCAACCGCTGAAGTGTCTGCTTCCGACCTTCCGAGGACCAAGGAGCTGG 355  
QY 1046 AGCTGTTCGCGAGGCGGTGCACCCCGCATTTAGGAGTGTGTACGA 1091  
Db 356 TGGGTTCTCGCGCTGGAAGGTACGCGCAAGTGGCCAGGCTTA 401

## RESULT 10

US-10-620-914-32  
Sequence 32, Application US/10620914

Publication No. US20040093639A1  
GENERAL INFORMATION:  
APPLICANT: Benning, Christoph  
APPLICANT: Riekhof, Wayne  
APPLICANT: Klug, Rouven  
TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids  
FILE REFERENCE: MSU-07769  
CURRENT APPLICATION NUMBER: US/10/620,914  
CURRENT FILING DATE: 2003-07-16  
PRIOR APPLICATION NUMBER: 10/118,495  
PRIOR FILING DATE: 2002-04-08  
NUMBER OF SEQ ID NOS: 52  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 32  
TYPE: DNA  
ORGANISM: Sinorhizobium meliloti  
US-10-620-914-32

Query Match 3.6%; Score 70; DB 7; Length 1251;  
Best Local Similarity 52.8%; Pred. No. 2.5e-08;  
Matches 151; Conservative 0; Mismatches 135; Indels 0; Gaps 0;  
QY 806 TGTTCGCGCCACCTTCTGTACACGAGTCTGTGGGAGGACCCCGAGCGGATATGGAGG 865  
Db 116 TGCTCTTTCCGAGCTCGTCTACCGCAGATCTGGGAGGACCGATTGTGACATGGAAG 175  
QY 866 TGATGAGATCAACCCCAAGACACGGTGTGACCTGTAGTACGCGGCTGCAATGCC 925  
Db 176 CGATGAGATCCGCTCCGAGACATCGGATCTGACGATCGGTTCCGCGGCTGCAACATGC 235  
QY 926 TGAACCTGTGTTGAGGCGGCGCGGAGTGTGTGCTGAGTGTGAGGAGTGTGGC 985  
Db 236 TGACCTATCTCTCCGCGAGCTTCGCCGATAGACGTGTGATCTCAACCCCATCA 295  
QY 986 CGCGCTTCTGAGCTGAAGAGTGGCCATTCAGCAGCTGAGTGTGAGGAGTGTGGC 1045  
Db 296 TCGCGCTCAACCGCTGAAGTGTCTGCTTCCGACCTTCCGAGGACCAAGGAGCTGG 355  
QY 1046 AGCTGTTCGCGAGGCGGTGCACCCCGCATTTAGGAGTGTGTACGA 1091  
Db 356 TGGGTTCTCGCGCTGGAAGGTACGCGCAAGTGGCCAGGCTTA 401

## RESULT 11

US-10-118-495-22  
Sequence 22, Application US/10118495  
Publication No. US20030074688A1  
GENERAL INFORMATION:  
APPLICANT: Benning, Christoph  
APPLICANT: Riekhof, Wayne  
APPLICANT: Klug, Rouven  
TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids  
FILE REFERENCE: MSU-06897  
CURRENT APPLICATION NUMBER: US/10/118,495  
CURRENT FILING DATE: 2002-04-08  
PRIOR APPLICATION NUMBER: 60/283,812  
PRIOR FILING DATE: 2001-04-13  
NUMBER OF SEQ ID NOS: 42  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 22  
TYPE: DNA  
ORGANISM: Mesorhizobium loti  
US-10-118-495-22

Query Match 3.6%; Score 69.6; DB 5; Length 1248;  
Best Local Similarity 45.0%; Pred. No. 3.1e-08;  
Matches 500; Conservative 0; Mismatches 579; Indels 33; Gaps 5;  
QY 704 TCCCTTACGTGCGGTGGTGGCGGCCCCCTACTACTGTGAGTGGCGCTGCCGCGG 763  
Db 11 TCTCTCGGATCTGGTGTTCGCGCGGCAAGGAAGTTGGAAGGCGGCTTACCAAGACC 70

```
QY 764 TTGGCCACGCGCTGCACGAGGCGCGTGGAGCGCGGCCCATGTTCCCGCCACCTTCC 823
Db 71 GCGCGCTTTCAAAGCGGATCTCCGAGCGCTGTTCGCTCTCTCTCTCTCTCTCTCG 130
QY 824 TGTACAGCGAGTCTGTGGAGGAGCCCGAGCGGATATGGAGGTGATGAGATCAACCCCA 883
Db 131 TCTATCCGAGATCTGGGAGAGCCCGATGTCACATGGAGGCGATGACAGTTGGTCAGG 190
QY 884 AGACACGGTGTGACCTGACCTAGCGCGGCTGCATGCCCTGAACTCTGTCGAGG 943
Db 191 GCCATGCAATCGTCACAATCGCTTCGCGCGGCTGCAACATCTCTGCTACCCGTT 250
QY 944 GCGCGCGCCAGGTGTGCTGGTGGACTGCAACCCCGCGAGTGGCGCTCTCTGAGCTGA 1003
Db 251 CGCGCGACGATCGAGCGCTGCACTCAACCGCGCCACATCGCGCTGAACCGCATGA 310
QY 1004 AGAAGTGGCCATTCAGCAGCTGAGGATTTGAGGAGCG---TGTGGCAGCTGTTCGGCGAGG 1060
Db 311 AGCTGGAGGCGGTGGCGCGTCTGCCCTCGCAGGCGCATCTGTTCCGCTTTTTCGGCGCG 370
QY 1061 GCGTGCACCGCGCATTTGAGGAGCTGTACGAGAAGAGCTGGCGCGCTTCTCTGCGCAA 1120
Db 371 CCGACACACGCCCAATTCGCAAGCCATATACCGCTTTATTTGGCGCGCATCTCGATCCGG 430
QY 1121 CCAGCCACAACTTCTGTGTCAAGCGCTCTGGTACTTCCAGCAGCGCTGTACTTACCAGG 1180
Db 431 TTAGCGCCCACTATTGGAGGCGCGCAACTTGGCGTGTTCGCGCGCATCGCCGTCTTCG 490
QY 1181 GCGGCATGGGCAAGCTGTGTGGGTGTCTGAGTGCCTTGGCGCTGTGGTGTGGGATCGGCA 1240
Db 491 ACCGCAATTTCTACGAGCC---GGCCTGTCTGCGCTGTTCATCGCATGGGCCATCGCA 547
QY 1241 AGACCGTCAAGCGCTCGCAACCGCGCCCAATGAGAGGAGCAGCGCGCTGTGGGACA 1300
Db 548 CGCGCAAAATTTCTGGCGGTCAACCGCGCCCAATGATGGAAGCCAGGAATATCGCGCAGC 607
QY 1301 GCAACATGCTCATCTCGTGAAGAAAGCGGCCCAAGCGCTGTGTGGCTGTTCGTCGA 1360
Db 608 AGCGCGCTTCTTCAACGAG---AGCTGGCGCGCTTTCGACAGA 652
QY 1361 AGTTCTGAGCGCTGTCTCTTCAACAGGCGGCTGTGTGTGGTGGCGGCGGTGCGCG 1420
Db 653 AGCTTTTGAATGGCGACCTCGGTAAAGCTCTGCTGTTCGGCTTCGGCATTCGCGCGG 712
QY 1421 GCAAGAGTAGCGCTGATCAAGCGGAGCGCATCCCGCATTTGAGAACTACATCGCGGCA 1480
Db 713 CGCAGTACGATTCCTGATCACTCAGCGCGCAGCGCACCATTTGGCAGCGTTCTGAAGGCC 772
QY 1481 CCATGGACCGGTGGCGGAGAACTCGCAGCTGCGCAAGCAGAACTACTTCTACTACAACT 1540
Db 773 GGCTGGAAAGCTCGCTCGATTTTCCCTGGAAACAAATTTTCCGCTGGCAGGCTT 832
QY 1541 GCCTCACCGGCAAGTTCTCTGCGGACAACTGC-----CCACCTACTCTGCGGAGCGG 1594
Db 833 TTGCGCGCGCTATCCAAATCCCGGTGAGCGCGCGCTTCCCGCTATCTGGAAAGCAGA 892
QY 1595 CTTTGGCCACCT-----CAAGAGTGGGTGGTGGACAACTGACCGTCTTCCACCACT 1648
Db 893 ACTACGAAACCATCCGCGGCAATATCGACCGCGTTCGCCATTCACCATGCAATCTGATCG 952
QY 1649 TCTTCATGGAGGAGCTCAAGCGGCGCACTTACCAAGGTGATTTCTGATGACCACTGG 1708
Db 953 AATTCTCTCGCGCAAGGACGCGGCGCACCGTCTGATCGTTCTCTCTCTGATCGCGAGG 1012
QY 1709 ACTGGCTGGATATCCCGTGGCCCAACGAGCTGGCGGAGTGTCTGGCCCAAGCAGTTTCGC 1768
Db 1013 ACTGATGACCGATGACAGCTCAACGCGCTGTGTTCTGGAATTCAGCGCGACCGCTCCG 1072
QY 1769 CGGCGCGCATGTGATCTGTGGCGCTTCGCGCTCC 1800
Db 1073 CAGCGCGCGCGTATCTTTCGCGCACCGCGCC 1104
```

```
RESULT 12
US-10-620-914-22
; Sequence 22, Application US/10620914
; Publication No. US20040093639A1
; GENERAL INFORMATION:
; APPLICANT: Benning, Christoph
; APPLICANT: Riekhof, Wayne
; APPLICANT: Klug, Rouven
; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids
; FILE REFERENCE: MSU-07769
; CURRENT APPLICATION NUMBER: US/10/620,914
; CURRENT FILING DATE: 2003-07-16
; PRIOR APPLICATION NUMBER: 10/118,495
; PRIOR FILING DATE: 2002-04-08
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 22
; TYPE: DNA
; ORGANISM: Mesorhizobium loti
; US-10-620-914-22

Query Match 3.6%; Score 69.6; DB 7; Length 1248;
Best Local Similarity 45.0%; Pred. No. 3.1e-08;
Matches 500; Conservative 0; Mismatches 579; Indels 33; Gaps 5;

QY 704 TCCCTACGTCGCTGCTGCGCGCCCTACTACGTGTGGATTGGCGCGCTGCCACGCG 763
Db 11 TCTCTCTCGATCTGGTTTTCGCGCGGCAAGAAAGTTGGAAGCGCGCTTACAGAAC 70
QY 764 TTGGCCACGCGCTGCACGAGGAGCGGTGGAGCGCGCGCCCATGTTCCCGCCACCTTCC 823
Db 71 GCGCGCTTTCAAAGCGGCACTCTCCGAGCGCTGTTCGCTTCTCTCTCTCTCTCG 130
QY 824 TGTACAGCGAGTCTGTGGAGGAGCCCGAGCGGATATGGAGGTGATGAGATCAACCCCA 883
Db 131 TCTATCCGAGATCTGGGAGAGCCCGATGTCACATGGAGGCGCATGACAGTTGGTCAGG 190
QY 884 AGACACGGTGTGACCTGACCTGAGCGCGCTGCAATGCCCTGAACTCTGTCGTGCGAG 943
Db 191 GCCATGCGATCTGTCACAATCGCTTCGCGCGGCTGTCAACATCTCTGCGCTACTTACCCTGT 250
QY 944 GCGCGCGCCAGGTGTGCTGGTGGAGTCTCAACCCCGCGAGTGGCGCTTCTGAGAGCTGA 1003
Db 251 CGCGCGACGATCGAGCGCTGCACTCAACCGCGCCACATCGCGCTGAACCGCATGA 310
QY 1004 AGAAGTGGCCATTCAGCAGCTGAGGATTTGAGGAGCG---TGTGGCAGCTGTTCGGCGAGG 1060
Db 311 AGCTGGAGGCGGTGGCGCGTCTGCCCTCGCAGGCGCATCTGTTCCGCTTTTTCGGCGCG 370
QY 1061 GCGTGCACCGCGCATTTGAGGAGCTGTACGAGAAGAGCTGGCGCGCTTCTCTGTCGCAA 1120
Db 371 CCGACACACGCCCAATTCGCAAGCCATATACCGCTTTATTTGGCGCGCATCTCGATCCGG 430
QY 1121 CCAGCCACAACTTCTGTGTCAAGCGCTCTGGTACTTCCAGCAGCGCTGTACTTACCAGG 1180
Db 431 TTAGCGCCCACTATTGGAGGCGCGCAACTTGGCGTGTTCGCGCGCATCGCCGTCTTCG 490
QY 1181 GCGGCATGGGCAAGCTGTGTGGGTGTCTGAGTGCCTTGGCGCTGTGGTGTGGGATCGGCA 1240
Db 491 ACCGCAATTTCTACGAGCC---GGCCTGTCTGCGCTGTTCATCGCATGGGCCATCGCA 547
QY 1241 AGACCGTCAAGCGCTCGCAACCGCGCCCAATGAGAGGAGCAGCGCGCTGTGGGACA 1300
Db 548 CGCGCAAAATTTCTGGCGGTCAACCGCGCCCAATGATGGAAGCCAGGAATATCGCGCAGC 607
QY 1301 GCAACATGCTCATCTCGTGAAGAAAGCGGCCCAAGCGCTGTGTGGCTGTTCGTCGA 1360
Db 608 AGCGCGCTTCTTCAACGAG---AGCTGGCGCGCTTTCGACAGA 652
QY 1361 AGTTCTGAGCGCTGTCTCTTCAACAGGCGGCTGTGTGTGGTGGCGGCGGTGCGCG 1420
Db 653 AGCTTTTGAATGGCGACCTCGGTAAAGCTCTGCTGTTCGGCTTCGGCATTCGCGCGG 712
QY 1421 GCAAGAGTAGCGCTGATCAAGCGGAGCGCATCCCGCATTTGAGAACTACATCGCGGCA 1480
Db 713 CGCAGTACGATTCCTGATCACTCAGCGCGCAGCGCACCATTTGGCAGCGTTCTGAAGGCC 772
QY 1481 CCATGGACCGGTGGCGGAGAACTCGCAGCTGCGCAAGCAGAACTACTTCTACTACAACT 1540
Db 773 GGCTGGAAAGCTCGCTCGATTTTCCCTGGAAACAAATTTTCCGCTGGCAGGCTT 832
QY 1541 GCCTCACCGGCAAGTTCTCTGCGGACAACTGC-----CCACCTACTCTGCGGAGCGG 1594
Db 833 TTGCGCGCGCTATCCAAATCCCGGTGAGCGCGCGCTTCCCGCTATCTGGAAAGCAGA 892
QY 1595 CTTTGGCCACCT-----CAAGAGTGGGTGGTGGACAACTGACCGTCTTCCACCACT 1648
Db 893 ACTACGAAACCATCCGCGGCAATATCGACCGCGTTCGCCATTCACCATGCAATCTGATCG 952
QY 1649 TCTTCATGGAGGAGCTCAAGCGGCGCACTTACCAAGGTGATTTCTGATGACCACTGG 1708
Db 953 AATTCTCTCGCGCAAGGACGCGGCGCACCGTCTGATCGTTCTCTCTCTGATCGCGAGG 1012
QY 1709 ACTGGCTGGATATCCCGTGGCCCAACGAGCTGGCGGAGTGTCTGGCCCAAGCAGTTTCGC 1768
Db 1013 ACTGATGACCGATGACAGCTCAACGCGCTGTGTTCTGGAATTCAGCGCGACCGCTCCG 1072
QY 1769 CGGCGCGCATGTGATCTGTGGCGCTTCGCGCTCC 1800
Db 1073 CAGCGCGCGCGTATCTTTCGCGCACCGCGCC 1104
```

Db 653 AGCTTTTGAATGGGCGACCTCGGTTAGGCTCGCTGTTGCTGGCTCGGCAATTCGCGCG 712  
Qy 1421 GCAAGCAGTACGGGCTGATCAAGCGGAGCGGATCCCATTTGAGAACTACATCGCGCGCA 1480  
Db 713 CGCAGTACGATTCCCTGATCACCTCAGGCGAGCGCACCATTGCGCAGGTTCTGAAGGCC 772  
Qy 1481 CATGAGCGGCTGGCGGAGAACTCGACGTGGCGAGCAGAACTACTTCTACTACAACT 1540  
Db 773 GGCTGAAAAGAGCTCGCTGCGATTCTTCCCTCGGAAACAAATTTATTCGCTGCGCAGGCTT 832  
Qy 1541 GCCTCACCAGGCAAGTTCTCGCGGCAAACTGC-----CCCACTACCTGGCGAGCGCG 1594  
Db 833 TTGCGCGCGCTATCCAAATCCCGGTGAGCGCGCCCTGCGCCCTATCTGGAAGAGAGA 892  
Qy 1595 CTTTCCGCAACCT-----CAAGAGTGGGTGGGAGCAACCTGACCGTCTCCACCAACT 1648  
Db 893 ACTACGAAACCATCCGCGGCAATATCGACCGGTGCGCATCCACCATGCAATCTGATCG 952  
Qy 1649 TCTTCATGAGGAGCTCAAGCGCGCACTACACCAAGGTGATTCGATGAGCAGCGTGG 1708  
Db 953 AATTCTCGCGGCAAGGACGCGGCAACCGTCGATCGCTTCACTCTGCTGATGCGCAGG 1012  
Qy 1709 ACTGGCTGGATATGCCGTGGCAACGAGCTGCGCGAGTGCCTGGCAAGCAGGTTGGCG 1768  
Db 1013 ACTGGATACCGATGACGACTCAAGCGCTGTGGTGGAAATCAGCGGACCGCCTCG 1072  
Qy 1769 CGGCGGCGCATCGTATCTGCGCTCGCGCTCC 1800  
Db 1073 CAGGCGCGCGTCACTCTTCCGCAACCGCGCC 1104

RESULT 13  
US-10-204-434A-12  
; Sequence 12, Application US/10204434A  
; Publication No. US20030170662A1  
; GENERAL INFORMATION:  
; APPLICANT: SETO, Haruo  
; APPLICANT: KUZUYAMA, Tomohisa  
; TITLE OF INVENTION: A Method for Screening A Substance Which Specifically Inhibits No  
; FILE OF INVENTION: Mevalonate Pathway  
; FILE REFERENCE: P22747  
; CURRENT APPLICATION NUMBER: US/10/204,434A  
; CURRENT FILING DATE: 2002-08-30  
; PRIOR APPLICATION NUMBER: PCT/JP01/01501  
; PRIOR FILING DATE: 2001-02-28  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 12  
; LENGTH: 1104  
; TYPE: DNA  
; ORGANISM: Streptomyces sp.  
US-10-204-434A-12

Query Match 3.5%; Score 67.8; DB 6; Length 1104;  
Best Local Similarity 45.1%; Pred. No. 9.3e-08;  
Matches 252; Conservative 0; Mismatches 307; Indels 0; Gaps 0;

Qy 1344 GGTGTGGCTGTTCTCGTCAAGTTCTGAGCTGCTCTTCAACAGCGCGTGTGTT 1403  
Db 513 GGTCCCGCAGATCGAGAGATCGCGGCGCGCTCGACATCCCGTGTATGTTCAAGAGGT 572  
Qy 1404 CGCGCGCGGCTCGCGGCAAGCAGTACGCGCTGTGATCAAGCGGAGCGGATCCCAATTGA 1463  
Db 573 CGGCAACGGCTGAGCGCGGAGACCATCTGCTGCTCGCGACCTCGGCGTGCAGCGCG 632  
Qy 1464 GAATACATCGCGCGCAACATGACGCGGTGGGAGAACTCGCAGTGGGCAAGCAGAA 1523  
Db 633 GGACGTGAGCGCGCGGCGGCAAGGCTTCCCGCATCGAGAACGCGCGCGGAGCT 692  
Qy 1524 CTACTTCTACTACATGCTCACCAGCAAGTTCTCTGCGCGCAAACTGCGCCCACTACCT 1583  
Db 693 CGCGGACTAGCGGTTCTGACGGCTGGGGGAGTGCACCGCGCGCTGCTGCTGAGACGC 752

Qy 1584 GCGCGAGCGCGCTTCCGCCACCTCAAGAGTGGCTGGTGGACAACCTGACCTCTCCAC 1643  
Db 753 CCAGGACATCTCCCTCCCGTCTCTCCCTCCCGCGTGTGCGTCAACCGCTCGACGTTGT 812  
Qy 1644 CAACTTCTCATGAGGAGCTCAAGCGGCGACCTACACCAAGGTGATTTCTGATGACCA 1703  
Db 813 CGCGCGCTTCCGCTCGCGCGCGCGCTCTCTCGCGCGCTCTCTCGCGCGCTTCTCGCGCACCT 872  
Qy 1704 CBTGGACTGGCTGGATATGCCGTGGCCCAACGAGCTGGCGGAGTGGCTGGCCAAAGCAGGT 1763  
Db 873 GATGGACGAGCGCTCGACCGCTGATCAGAGCTCAGACCTGGCTGGACGAGCTGGC 932  
Qy 1764 TGGCGCGCGGCGGATCGTATCTGGCGCTCGCGCTCTCTCAGCGCGCGCTTACGCGAGCT 1823  
Db 933 GCGCTTGCAGACCATCTCTCGCGCGCGCACCCCGCGCGACCTCAGCGCTGCGACGTTGT 992  
Qy 1824 GATCCAGAGGCGGCTTTCGAGTGGCTGATCCCGCGCGGCGCACCTCAGGGCTACATGGA 1883  
Db 993 GCTTCCACGCGGAGCTCGGTGACTTCTGCGCGGACCGGCGGATCGACACGCGCGCTCGC 1052  
Qy 1884 CCGCGTCAACATGTATACAGC 1902  
Db 1053 CCAGCGCTCCAGCTCCATC 1071

RESULT 14  
US-09-918-740-57  
; Sequence 57, Application US/09918740  
; Publication No. US20030033626A1  
; GENERAL INFORMATION:  
; APPLICANT: Hahn, Frederick  
; APPLICANT: Kuehnle, Adelheid  
; TITLE OF INVENTION: Manipulation of genes of the mevalonate and isoprenoid pathways t  
; FILE OF INVENTION: create novel traits in transgenic organisms  
; FILE REFERENCE: KAS-103XC1  
; CURRENT APPLICATION NUMBER: US/09/918,740  
; CURRENT FILING DATE: 2001-07-31  
; PRIOR APPLICATION NUMBER: 60/221,703  
; PRIOR FILING DATE: 2000-07-31  
; NUMBER OF SEQ ID NOS: 76  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 57  
; LENGTH: 6798  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Streptomyces sp CL190 gene cluster containing mevalonate pathway  
; OTHER INFORMATION: and IPP isomerase orfs  
US-09-918-740-57

Query Match 3.5%; Score 67.8; DB 3; Length 6798;  
Best Local Similarity 45.1%; Pred. No. 1.4e-07;  
Matches 252; Conservative 0; Mismatches 307; Indels 0; Gaps 0;

Qy 1344 GGTGTGGCTGTTCTCGTCAAGTTCTGAGCTGCTCTTCAACAGCGCGTGTGTT 1403  
Db 3835 GGTCCCGCAGATCGAGAGATCGCGCGCGCTCGACATCCCGTGTATGTTCAAGAGGT 3894  
Qy 1404 CGCGCGCGGCTCGCGGCAAGCAGTACGCGCTGTATCAAGCGGAGCGGATCCCAATTGA 1463  
Db 3895 CGGCAACGGCTGAGCGCGGAGACCATCTGCTGCTGCGCGACCTCGGCGTGCAGCGGC 3954  
Qy 1464 GAATACATCGCGCGCAACATGAGACGCGGTGGCGAGAACTCGCACTGCGCAAGCAGAA 1523  
Db 3955 GGACGTGAGCGCGCGCGGCGGCGGACGACCTTCGCGCATCGAGAACGCGCGCGGAGCT 4014  
Qy 1524 CTACTTCTACTACAACTGCTCAGCGGCAAGTTCTGCGCGGCAAACTGCGCCCACTACCT 1583  
Db 4015 CGCGGACTAGCGGTTCTCTGACGGCTGGGGGAGTGCACCGCGCGCTGCTGCTGAGACGC 4074  
Qy 1584 GCGCGAGGCGGCTTCCGCCACCTCAAGAGTGGCTGGTGGACAACCTGACCTCTCCAC 1643  
Db 4075 CCAGGACATCTCTCTGCGCGCTCTCTGCGCTCCCGCGGTGTGCGTCAACCGCTCGACGTTGT 4134

QY 1644 CAACTTCTTCATGAGGAGCTCAAAAGCGCGCACCTACACCAAGGTGATTCGTATGACCA 1703  
Db 4135 CCGGCGCCTCGCGTCCGCGCCCGCGCTCGGCTCTCCGCGGCTTCCTGCGCACCT 4194  
QY 1704 CGTGGACTGCTGATATGATCCGCTGGGCAACGAGCTGGCGAGTGGCTGGCCCAAGCAGGT 1763  
Db 4195 GATGGACGACGGCGTGCACGGCTGATCAGAGCTCACAGCTGGCTGGACCAAGCTGGC 4254  
QY 1764 TGGCGCGGGCGGATCGTATCTGGGCTCGGCTCCCTCAGCCCGGCTTACGCCGAGCT 1823  
Db 4255 GCGCTGACAGACCATGCTCGCGCGCGCACCCGCGCGAGCTCACCGCTGCGACGCTGCT 4314  
QY 1824 GATCCAGAAGCGGGCTTCAGCTGGCTGCATCCGCGCGCCACTCAGGGCTTACATGGA 1883  
Db 4315 GCTCCAGCGGAGCTGCGTGAATCTTCCGCGCGACCGGGGATCGACACGCGCGCTCGC 4374  
QY 1884 CCGCGTCAACATGTACAGC 1902  
Db 4375 CCAGCGTCCAGCTCCATC 4393

## RESULT 15

US-10-204-434A-13  
; Sequence 13, Application US/10204434A  
; Publication No. US20030170662A1  
; GENERAL INFORMATION:  
; APPLICANT: SETO, Haruo  
; APPLICANT: KUZUYAMA, Tomohisa  
; TITLE OF INVENTION: A Method For Screening A Substance Which Specifically Inhibits No  
; TITLE OF INVENTION: Mevalonate Pathway  
; FILE REFERENCE: P22747  
; CURRENT APPLICATION NUMBER: US/10/204,434A  
; CURRENT FILING DATE: 2002-08-30  
; PRIOR APPLICATION NUMBER: PCT/JP01/01501  
; PRIOR FILING DATE: 2001-02-28  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 13  
; LENGTH: 6798  
; TYPE: DNA  
; ORGANISM: Streptomyces sp.  
US-10-204-434A-13

Query Match 3.5%; Score 67.8; DB 6; Length 6798;  
Best Local Similarity 45.1%; Pred. No. 1.4e-07;  
Matches 252; Conservative 0; Mismatches 307; Indels 0; Gaps 0;  
QY 1344 GGTGTGCTGTTCGTCAAGTTTCGTGAGCTTGTGCTCTTCAACAAGCGCTGTGTGTT 1403  
Db 3835 GGTCCCGAGATCAGAGATCGCGCGCGCTGCATCCCGTGTATCGTCAAGGAGT 3894  
QY 1404 CGCGCGCGGTGCGCGGCAAGCATAGCTGCTGATCAAGCGGACCGCATCCCATTTGA 1463  
Db 3895 CGGCAACGGCTGAGCGCGGACCATCTGCTCTCGCGACCTCGCGCTGCGGCGC 3954  
QY 1464 GAATCATATCGCGCGACCATGAGCGGTGGCGAGACTCGACATGCGGCGAGCAAA 1523  
Db 3955 GGACGTGAGCGCGCGCGCGCACGGACTTGGCCGATCGAGAACCGCGCGGAGCT 4014  
QY 1524 CTACTTCTACTACAACTGCCTCACCGCAAGTTCCTGCGGACAACTGCGCCACCTACCT 1583  
Db 4015 CGGCGACTACCGGTTCTGACGGCTGGGGGAGTCCACCGCGCTTCCCTGTGGAGCG 4074  
QY 1584 GCGCGAGCGGCTTCGCGACCTCAAGAGTGGTGGTGGACAACTGACCGCTCTCCAC 1643  
Db 4075 CCAGGACATCTCCCTGCGCGCTCTCGCTCGCGGCTGTGCTCACCGCTCGACGCTGT 4134  
QY 1644 CAACTTCTCATGAGGAGCTCAAGCGCGCACCTACACCAAGGTGATTCGTATGACCA 1703  
Db 4135 CCGCGCTTCGCGTCCGCGCGCGCGCGCTCGGCTCTCCGCGGCTTCTCGCGCACCT 4194  
QY 1704 CGTGGACTGGCTGGATATATCCCGTGGCGCAACGAGCTGGCGGCTGCGTGGCCAAAGCAGGT 1763

Search completed: March 14, 2006, 04:22:33  
Job time : 1630 secs



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 14, 2006, 01:11:21 ; Search time 659 Seconds  
(without alignments)  
6816.854 Million cell updates/sec

Title: US-10-620-914-44  
Perfect score: 1947  
Sequence: 1 atgggggtcgggtcgtgacgg.....gcgccaaagaagacaactaa 1947

Scoring table: IDENTITY NUC

Gapop 10\_0 , Gapext 1.0

Searched: 7673375 seqs, 1153648444 residues

Total number of hits satisfying chosen parameters: 15346750

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA\_New.\*

- 1: /cgn2\_6/prodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 2: /cgn2\_6/prodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/prodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/prodata/1/pubpna/PCT\_NEW\_PUB.seq.\*
- 5: /cgn2\_6/prodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/prodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 7: /cgn2\_6/prodata/1/pubpna/US10\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/prodata/1/pubpna/US10\_NEW\_PUB.seq.\*
- 9: /cgn2\_6/prodata/1/pubpna/US11\_NEW\_PUB.seq.\*
- 10: /cgn2\_6/prodata/1/pubpna/US11\_NEW\_PUB.seq.\*
- 11: /cgn2\_6/prodata/1/pubpna/US11\_NEW\_PUB.seq.\*
- 12: /cgn2\_6/prodata/1/pubpna/US11\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/prodata/1/pubpna/US60\_NEW\_PUB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	66	3.4	1577	9	US-11-096-568A-10906
2	64	3.3	2736	8	US-10-858-730-38
3	60.8	3.1	1325	9	US-11-096-568A-1077
4	59.6	3.1	1347	9	US-11-096-568A-12531
5	57.8	3.0	1599	9	US-11-096-568A-26187
6	57.6	3.0	1444	9	US-11-096-568A-20988
7	57.2	2.9	1306	12	US-11-166-609-1
8	56.2	2.9	3513	8	US-10-858-730-142
9	56	2.9	4282	12	US-11-183-624-1
10	56	2.9	4282	12	US-11-183-624-3
11	55.2	2.8	1243	9	US-11-096-568A-25531
12	55.2	2.8	88421	12	US-11-205-109-1
13	54.8	2.8	88421	12	US-11-205-109-1
14	54.6	2.8	1283	9	US-11-096-568A-23551
15	54	2.8	2025	12	US-11-167-048-2
16	53.8	2.8	2403	7	US-10-649-457-4
17	53.8	2.8	37507	8	US-10-522-037-2
18	53.6	2.8	1182	9	US-11-096-568A-26293
19	53.6	2.8	1580	9	US-11-096-568A-20206
20	53.6	2.8	2873	12	US-11-136-527-2776

ALIGNMENTS

RESULT 1

US-11-096-568A-10906  
; Publication 10906, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; FILE OF INVENTION: Therby  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 10906  
; LENGTH: 1577  
; TYPE: DNA  
; ORGANISM: Triticum aestivum  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(1577)  
; OTHER INFORMATION: Ceres Seq. ID no. 13597454  
US-11-096-568A-10906

Query Match		3.4%	Score 66;	DB 9;	Length 1577;
Best Local Similarity		47.2%	Pred. No. 2.5e-06;		
Matches 201;		Conservative 0;	Mismatches 225;	Indels 0;	Gaps 0;
QY	1408	GGCGGCGTCCCGGCAAGCAGTACGCTGATCAAGCGCGGCGCATCCCATTTGAGAAC	1467		
DB	315	GGATACATGCGCGGCGGCTCTACGACTGGACGGCTCCAGTACGGCAGCGGGTGGAG	374		
QY	1468	TACATGCGCGCACCATTGACGCGCTGGCGGAGAACTCCGACGTGCGCAAGCAGAACTAC	1527		
DB	375	CTCAGGTGCTCATCGCGGCTTCGGTGCCAAAGGCGATCAAGTGGCTGCCGACATCGTC	434		
QY	1528	TTCTACTCAACTGCTTCCCGGCAAGTTCCTGCGGGGCAACTGCGCCACCTACCTCGGC	1587		
DB	435	ATCAACACCGCTGCGCGGAGCAAGGACGGCGCGGCTTCTGCTCTTCAAGGGC	494		
QY	1588	GAGGCGGCTTCCGACACCTCAAGAGTGGCTGGTGGAACAACCTGACCTCTCCACCAAC	1647		
DB	495	GGAGGACCGAGGGGCTCTCTGACTGGGTGCGGGCATGATCTGACGACGACACCAAG	554		
QY	1648	TTCTTTCATGAGGAGTCAAGCGGCACCTACACCAAGGTGATTTCTGATGACCAACGCTG	1707		

Sequence 19003, A  
Sequence 26982, A  
Sequence 183, App  
Sequence 851, App  
Sequence 19140, A  
Sequence 25607, A  
Sequence 631, App  
Sequence 22550, A  
Sequence 6, Appli  
Sequence 20876, A  
Sequence 55, Appl  
Sequence 8, Appli  
Sequence 7, Appli  
Sequence 2, Appli  
Sequence 4, Appli  
Sequence 48926, A  
Sequence 48926, A  
Sequence 1, Appli  
Sequence 13310, A  
Sequence 58, Appl  
Sequence 23028, A  
Sequence 17222, A  
Sequence 21800, A  
Sequence 14592, A  
Sequence 19863, A

Db 555 TTCTCCGACGCGCGGTACCTGATACCGCGCGGACTTCGCGCGCGCGGACATC 614  
Qy 1708 GACTGGCTGATATGCCCCGTCGACGAGCTGGCGAGTGCTGGCCAAAGAGGTGG 1767  
Db 615 GACCACCTCAACCGCGCGGTCCAGAGAGAGCTCTCCGACTGGCTCAAGTCC 674  
Qy 1768 CCGGGCGGATGTCATCTGCGCTCCGCTCCCTCAGCCCGCCCTACGCGGAGCTGATC 1827  
Db 675 GACCTGGCTTCAGCGCTGGCGCTGACTTCGCCAAGGGCTACTCGGCCGAGCTGCC 734  
Qy 1828 CAGAAG 1833  
Db 735 AGGACG 740

## RESULT 2

US-10-858-730-38  
; Sequence 38, Application US/10858730  
; Publication No. US20050255568A1  
; GENERAL INFORMATION:  
; APPLICANT: Bailey, Richard B.  
; APPLICANT: Blomquist, Paul  
; APPLICANT: Doten, Reed  
; APPLICANT: Driggers, Edward M.  
; APPLICANT: Madden, Kevin T.  
; APPLICANT: O'Leary, Jessica  
; APPLICANT: O'Toole, George  
; APPLICANT: Trueheart, Joshua  
; APPLICANT: Walbridge, Michael J.  
; APPLICANT: Yorgey, Peter S.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR AMINO ACID  
; TITLE OF INVENTION: PRODUCTION  
; FILE REFERENCE: 14184-030001  
; CURRENT APPLICATION NUMBER: US/10/858,730  
; CURRENT FILING DATE: 2004-06-01  
; PRIOR APPLICATION NUMBER: US 60/475,000  
; PRIOR FILING DATE: 2003-05-30  
; PRIOR APPLICATION NUMBER: US 60/551,860  
; PRIOR FILING DATE: 2004-03-10  
; NUMBER OF SEQ ID NOS: 364  
; SOFTWARE: Fast-Seq for Windows Version 4.0  
; SEQ ID NO 38  
; LENGTH: 2736  
; TYPE: DNA  
; ORGANISM: Streptomyces coelicolor  
US-10-858-730-38

Query Match 3.3%; Score 64; DB 8; Length 2736;  
Best Local Similarity 47.6%; Pred. No. 7.6e-06;  
Matches 221; Conservative 0; Mismatches 240; Indels 3; Gaps 1;  
Qy 1458 CATTGAACTATCATCGCGCGCACCATGAGACGGGTGGCGGAGAACTCGCACTGCGCAA 1517  
Db 1011 CACCAAGCAGCGCTCCCAAGGCGACCCGCCAGAGAGCGCGCGCACTACCTCGGCAC 1070  
Qy 1518 GCAGAACTACTTCTACTACAACTCCCTCACGGCAAGTCTTCGGCGACAACTGCCCCAC 1577  
Db 1071 CGCCACGCTCATGACACCTGGCGCATCTGTCAGACCTCGTGGCGAAACACCGCGCGG 1130  
Qy 1578 CTACCTCGCGAGCGCGCTTCGCCACCTCAAGAGTGGCGGTGTGG---ACAACCTGAC 1634  
Db 1131 CTGTTTCGCGACGGGGCGCTTCGCCCGCACCATCGGCACCTTGGCGCGCTTCGCCCTCA 1190  
Qy 1635 CGTCTCCACCACTTCTTTCATGAGAGAGCTCAAAGCGCGCACTTACACCAAGGTGATCT 1694  
Db 1191 GCTCGCACCATGAGCTTCGCGAGACGCGCGAGCCACCAACACCGCCCTCGGCCAGCT 1250  
Qy 1695 GATGACCACTGAGCTGGCTGGATATGCCGTGGCGCAAGAGTGGCCGAGTGCCTGGC 1754  
Db 1251 CTTGACCCGGCTCGCGAGAGATCTCTGGCGCTACGCCGATGCGCGGAGTACCGCAC 1310  
Qy 1755 CAAGACGTTTGGCGCGCGCGCATCTGTCATCTGGCGCTCGCGCTCCCTCGACCGCCCTA 1814

Db 1311 CRAAGTCTCTCGCCAGGAACCTGGCTCCCGCAGGCGCTGGCCCCCAGCCCCCGCCGT 1370  
Qy 1815 CGCGAGCTGATTCACAAAGCGGGCTTCGACGTGGCTGCCTGCATCGCGGCGGCGCACTCAGGG 1874  
Db 1371 CGACGCGCCCGCGGAGAACCTCTCGCGCTTTTCCAGACCGTCCGCGCGCCCTGGAGGT 1430  
Qy 1875 CTACATGGACCGGTCAACATGTACAGCTCTCTTCTACATGGCCC 1918  
Db 1431 CTTGCGCCCGGAGTTCATCGAGTCTTACATCATCTCCATGTGCC 1474

## RESULT 3

US-11-096-568A-1077  
; Sequence 1077, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; TITLE OF INVENTION: Therby  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 1077  
; LENGTH: 1325  
; TYPE: DNA  
; ORGANISM: Zea mays subsp. mays  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(1325)  
; OTHER INFORMATION: Ceres Seq. ID no. 13599631  
US-11-096-568A-1077

Query Match 3.1%; Score 60.8; DB 9; Length 1325;  
Best Local Similarity 45.4%; Pred. No. 4.2e-05;  
Matches 258; Conservative 0; Mismatches 307; Indels 3; Gaps 1;  
Qy 1218 GCGCGTGTGTCTGGGACTGGGCAAGACCGTCAAGCGCTCGCCAAACGCGCCCAATGGA 1277  
Db 293 GCGGTGGGTTCGGCAAGCCAGCCGCGGAGCTGGCCGATCATGTGCCACATCGG 352  
Qy 1278 GAGCAGCGCCGTCTGTGGGACAGCAACATGCTCATCTCTGTGAAGAACCGGCCAA 1337  
Db 353 CTTCCACCGCGCGACTCTGTCTGTGACGTCTCATCAAGAAACCCCAACCGGTGCCAT 412  
Qy 1338 GCGCTGTGTGTGTCTGTCTCAAGTTCGTGAGCTGTGTCTCTTCAACAGGCGGTCT 1397  
Db 413 CCGCTCGTGCACATCGACTTACTCTATCGACAGCGCGCGGAAAGCTCTCGCGGGCT 472  
Qy 1398 GTGGTTTCGGCGCGGTGCGCGGCAAGCAGTACGCGCTGA---TCAAGCGGACGCGCAT 1454  
Db 473 CATCCGGACCGCGGACCATCTCGCGGCAAGCGGAGGAGACGCTCAAGTGGCCATCAC 532  
Qy 1455 CCCATTGAGAACTATCATCGCGCGCACCATGGAAGCGGTGGCGGAGAACTGCACTGGG 1514  
Db 533 GCTCGACTTCGACGACATCGGAGCACCTACGCGGACATCAAGCGCGGAGCATATCCC 592  
Qy 1515 CAAGCAGAACTACTTCTTACTAATGCTCTCAAGCGGCAAGTCTCTGCGCGCAACTGCC 1574  
Db 593 CTACCTGCTCGGGTCTATCTTCTCTGCGGAGCTGCGCGCTCTTGGCGCGCATCAAGATCCC 652  
Qy 1575 CACTTACCTCGGAGCGCGCTTCGCCACCTTCAAGAGTGGGTGGTGGCAACCTGAC 1634  
Db 653 GCTGGACAACTCGGGAGATCTCCCATCCGTAACAGCCGACGTGACGTGGACAGAT 712  
Qy 1635 CGTCTCCACCACTTCTTCTATGAGAGAGCTCAAAGCGCGCACTTACACCAAGGTGATCT 1694  
Db 713 CAAGTTCCACCACTTCTTCTCTCGAGGAGACACGCGGCAACCATCCACCTGAGCTGGAGAA 772  
Qy 1695 GATGACCACTGAGCTGGCTGGATATGCCGTGGCGCAAGAGTGGCCGCGGAGTGCCTGGC 1754  
Db 773 CAAGAACGACTTCGACCTCGGGCTCAACCTGCTCAGTACGAGATGTGGTTCGCGGACGA 832



; TYPE: DNA  
; ORGANISM: Zea mays subsp. mays  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(1444)  
; OTHER INFORMATION: Ceres Seq. ID no. 12391561  
US-11-096-568A-20988

Query Match 3.08; Score 57.6; DB 9; Length 1444;  
Best Local Similarity 46.28; Pred. No. 0.00024; Indels 0; Gaps 0;  
Matches 192; Conservative 0; Mismatches 224; Indels 0; Gaps 0;

QY 1527 CTTCTACTACAACTGCTCCACCGGCAAGTTCTCGCGCAAACTGCCCCACCTACTCTGCG 1586  
DB 510 CTTCTGGCCCTACTGCTCTGAGTTCTGATGTGATATCTTCCCGGCACTCTCTGCGG 569  
QY 1587 CGAGGGGCGCTTCCCAACCTCAAGAGTGGGTGGCAAACTGACCTCTCTCCACCAA 1646  
DB 570 CGACCCGCGCTTACCGCGCGGCAAGCAACACCTTCTTCCATGGAAACAAGGACAA 629  
QY 1647 CTTCTTCTCATGAGAGCTCAAGCGGCACCTACCAAGGTGATTTCTGATGGACCACT 1706  
DB 630 GGAATCTTCTGCTCTTCCGACGCGGCGCTTCCACATCAACGCCCACTTCTATCGGCAAC 689  
QY 1707 GGAATGCTGATATGCTGCGGCAAGAGCTGGCGGAGTGGCTTGGCCAAAGCAAGTTGC 1766  
DB 690 CAACCCGCGCTTCAACCGGCACTTCAAGTGGTGGAGCGCTGGGGTCACTTCTGCTGC 749  
QY 1767 GCGGGGCGGATGCTCATCTGCGGCTTCCGCTTCCAGCGCGCTTCAAGCGGCTGAT 1826  
DB 750 ACCAGGCGGCGCGGCGGCAAGCAACCTTCTTACGTCGGCGCGCGGAGTG 809  
QY 1827 CCGAAGGCGGCTTCAAGTGGCTGCTATCCGCGCGGCACTCAGGCTTACATGACCG 1886  
DB 810 GAGAGGAGGAGGACGACCACTTCCAGCTTCACTTGGAGCGCGGAGCGCTGCGACG 869  
QY 1887 CGTCAACATGTACAGCTCTTCTACATGCGCGCGGCGGAGCGGCGGCAAGGAGCA 1942  
DB 870 CGTCAGAACGCGGCTGGGCGTCCAGGCGCTTCCGCGGCTCTTCCGTCAGGCGCA 925

RESULT 7  
US-11-166-609-1  
; Sequence 1, Application US/11166609  
; Publication No. US20060015968A1  
; GENERAL INFORMATION:  
; APPLICANT: ALBERTSEN, MARC C.  
; APPLICANT: FOX, TIM  
; APPLICANT: HUFFMAN, GARY  
; APPLICANT: TRIMNELL, MARY  
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES MEDIATING MALE FERTILITY AND  
; TITLE OF INVENTION: METHOD OF USING SAME  
; FILE REFERENCE: 1148CR  
; CURRENT APPLICATION NUMBER: US/11/166,609  
; CURRENT FILING DATE: 2005-06-24  
; PRIOR APPLICATION NUMBER: 10/412,000  
; PRIOR FILING DATE: 2003-04-11  
; PRIOR APPLICATION NUMBER: 09/670,153  
; PRIOR FILING DATE: 2000-09-26  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: PatentIn Ver. 3.3  
; SEQ ID NO 1  
; LENGTH: 1906  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1638)  
US-11-166-609-1

Query Match 2.98; Score 57.2; DB 12; Length 1906;  
Best Local Similarity 43.18; Pred. No. 0.00031;  
Matches 275; Conservative 0; Mismatches 363; Indels 0; Gaps 0;

QY 908 GCGGCGGCTGCAATGCTGCTGAACCTGCTGAGGGGGCGGCCAGAGTGTGCTGCGTGG 967  
DB 911 GCGGCGGCTTGGGGAGGATAGAGCTTCCGGGAGCTGGTGTCTCAACTTGTGATGCGCG 970  
QY 968 ACTGCAACCCCGCGAGTGGCGCTCTTCTGAGCTGAAGAAGTGGCCATTCAGAGCTGG 1027  
DB 971 GCGGCGACACGACGCGGACGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1030  
QY 1028 AGTTTGAAGAGCTGTCGCGAGCTGTTTGGCGGAGGCGCTGACCCGCGCATTCAGAGCTGT 1087  
DB 1031 ACGTGGCCGAGAGCTGCGCGCGAGCTGTGCGCGGTTCGAGGCGGAGCGCGCGCGAGG 1090  
QY 1088 ACCAGAGAAGCTGGGCGGCTTCTCTGCTGCAAAACAGCCACAACTTCTGCTTCCAAGCGCC 1147  
DB 1091 AGGGCGTCACTGCTGCTGCTGCGCGGCGCTGACGCGGACGACAGGCGCTTCGCGCGCC 1150  
QY 1148 TCTGCTGCTTCCAGCAGCGCTTCTATACAGGCGGCGATGGGCAAGCTGCTGCTGGTGC 1207  
DB 1151 GCGTGGCGAGTTTGGCGGCGCTTCTTCACTTACGACAGCTTGGGAGCTGCTTACCTCC 1210  
QY 1208 TGCAGTGGCTGGCGGCTGCTGGGAGTGGGCAAGCCGTCAGCGCTCGCCAAACGCGC 1267  
DB 1211 ACGCTGCGTACCGAGAGCTCCGCTGTACCCCGCGCTTCCCTCAGGACCCCAAGGGA 1270  
QY 1268 CCACAAATGGAGGAGCAGCGCGCTCTGTGGGACAGCAACATGCTCATCTCTGTAAGA 1327  
DB 1271 TCTGAGGAGCAGCTGCTGCGGACGGAAGGTGAGGCGCGGCGGATGCTGACGT 1330  
QY 1328 ACGGCGCCAAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTTCAACA 1387  
DB 1331 ACGTGGCTTACTGATGGGCGGATGAGTACAACTGGGCGCGCGGCGGAGCTTCC 1390  
QY 1388 AGCGCTGCTGCTGCTGCTGCGGCGGCTGCGGCGGAGGAGTACGCGCTCATCAAGCGG 1447  
DB 1391 GCGGCGAGCGGTGATCAACAGAGATGGCGCTTCCGCAACGCGTCCCGTTCAGTTCA 1450  
QY 1448 ACGGCAATCCCAATGAGAACTACATCGCGCGCACCATGAGCGGCTGGCGGAGAACTCGC 1507  
DB 1451 CGCGTTCCAGGCGGCGCGGAGGATCTGCTGGGCAAGGACTCGGCGTACTCGAGATGA 1510  
QY 1508 ACGTGGCGAAGCAGAACTACTTCTTACAACTGGCTC 1545  
DB 1511 AGATGGCGCTGGCATCTCTTCCGCTTCTACAGCTTC 1548

RESULT 8  
US-10-858-730-142  
; Sequence 142, Application US/10858730  
; Publication No. US20050255568A1  
; GENERAL INFORMATION:  
; APPLICANT: Bailey, Richard B.  
; APPLICANT: Blomquist, Paul  
; APPLICANT: Doten, Reed  
; APPLICANT: Driggers, Edward M.  
; APPLICANT: Madden, Kevin T.  
; APPLICANT: O'Leary, Jessica  
; APPLICANT: O'Toole, George  
; APPLICANT: Trueheart, Joshua  
; APPLICANT: Walbridge, Michael J.  
; APPLICANT: Yorgey, Peter S.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR AMINO ACID  
; TITLE OF INVENTION: PRODUCTION  
; FILE REFERENCE: 14184-030001  
; CURRENT APPLICATION NUMBER: US/10/858,730  
; CURRENT FILING DATE: 2004-06-01  
; PRIOR APPLICATION NUMBER: US 60/475,000  
; PRIOR FILING DATE: 2003-05-30  
; PRIOR APPLICATION NUMBER: US 60/551,860  
; PRIOR FILING DATE: 2004-03-10  
; NUMBER OF SEQ ID NOS: 364  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 142

; LENGTH: 3513  
; TYPE: DNA  
; ORGANISM: Streptomyces coelicolor  
US-10-858-730-142

```
Query Match      2.9%; Score 56.2; DB 8; Length 3513;
Best Local Similarity 44.6%; Pred. No. 0.00055;
Matches 266; Conservative 0; Mismatches 328; Indels 3; Gaps 1;

QY 1146 CCTCTGTACTTCCAGCACCGCTGTACTACAGGCGGATGGCAAGCTGTCTGGGT 1205
DB 1179 CTTCTGGCTGACTACGTCTGGCGGGAGCGGCTGCCGACATGAGGAACTTGGCCGGCG 1238
QY 1206 GCTGCACTGCTGCGCTGTGTGGACTGGCAAGACCGTCAAGCGCTCGCCCAACGC 1265
DB 1239 GTTCCGACCGCTCCACGCTGCGGCTCTCTGACTCCACGAGGTGCACTCCG 1298
QY 1266 GCCACAAATGAGGAGCAGCGCTGTGTGGACAGCAATGCTCATCTCTGTGAA 1325
DB 1299 GCGCGGCTGGAGAACTCGCGCGCGCGCGGTGATCAACTCGGTCAACTACGAGGACGG 1358
QY 1326 GAACGGGCCCAAGCGCTGTGTGGCTGTTCGTCAGTTCGTGAGCTGTGTCTTCAA 1385
DB 1359 CGCGGCGCCGAGTCCCGGTTGCGCGGTACGAACTGCGCGGAGCAGCGCGCGC 1418
QY 1386 CAAGGCGCTGTGTGTTCGCGCGCGCGCGTCCGCGCAAGCAGTACGCGCTGATCAAGGC 1445
DB 1419 GCTGATCGCGCTGACCATCAGCAGGTGGGACAGCGCGCCGCGGAGAAGAGTCTGA 1478
QY 1446 GAGCGGATCCCATTTGAGAACTACATCGCGCGCACATGAGCGGTGGCGGAGAACTC 1505
DB 1479 GATCGCGCAACGGCTCATCGACGACCTCACCGGCACTGGGGGATCCACGAGTCCGACAT 1538
QY 1506 GCAGCTGCGCAAGCAGAACTACTTCTACTAACTGCTCACCGGCAAGTTCTTGGCGGA 1565
DB 1539 CTTGTGCACTGCTGACCTTACCATCTGACCGGCTTACGCGGCTCCGCAAGGA 1595
QY 1566 CAACCTGCCCCACCTTACCTGCGGAGGCGGCTTTCGCCACCTCAAGAGTGGCGTGTGGA 1625
DB 1596 CGGCTTGGCCACCATCAGGAGGATCGGGAACCTCAAGCGCGCCACCGGAGCTGCAGAC 1655
QY 1626 CAACCTGACGCTTCCACCAACTTCTTCATGAGGAGTCAAGCGCGGACCTACACCAA 1685
DB 1656 CACGCTGGCGCTGTGCAACATCTCTTTCGGGCTCAACCGCGCGCGCGCATCTCTGTCTCA 1715
QY 1686 GGTGATTTCTGATGACCACTGACGTGCTGCTGATATGCGCTGCGCAAGAGCTGCG 1742
DB 1716 CTCGGTCTTCTCGACGAATGCGTCAAGCGCGGCTGGAATCGGCCCATCTGTGCACGC 1772
```

```
RESULT 9
US-11-183-624-1
; Sequence 1, Application US/11183624
; Publication No. US20050255518A1
; GENERAL INFORMATION:
; APPLICANT: Silos-Santiago, Immaculada
; TITLE OF INVENTION: Methods and compositions for the
; FILE OF INVENTION: treatment and diagnosis of pain disorders using 46556
; FILE REFERENCE: MPI01-272PIRM
; CURRENT APPLICATION NUMBER: US/11/183,624
; CURRENT FILING DATE: 2005-07-18
; PRIOR APPLICATION NUMBER: US/10/281,866
; PRIOR FILING DATE: 2002-10-28
; PRIOR APPLICATION NUMBER: 60/335,078
; PRIOR FILING DATE: 2001-10-31
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 4282
; TYPE: DNA
; ORGANISM: Homo sapien
US-11-183-624-1
```

```
Query Match      2.9%; Score 56; DB 12; Length 4282;
Best Local Similarity 44.2%; Pred. No. 0.00062;
Matches 282; Conservative 0; Mismatches 350; Indels 6; Gaps 1;

QY 1024 CTGGAGTTTGGAGACGCTGTGGCAGCTGTTCGGGAGGGCGTGACCCGCGCATTTGAGGAG 1083
DB 766 CAGGTGTGGAGGCGCTGTGTGACCCCTGGTCTTCTCCCGGTGTGCGTGTGATTGCGCTGG 825
QY 1084 CTGTACGAGAAAGCTGGCGCCCTTCTCTGTCGCAAAACAGCAGCACAACCTTCTGTGTCCAAG 1143
DB 826 ATGGCCGACAAAGCGGTGTCTTCTACAGTAGTGTACAGCGCTACCGCAGCGACCA 885
QY 1144 CGCTCTGTGTATTCACAGACGCGCTGTACTACAGGGCGGATGGCAAGTGTGTGTGG 1203
DB 886 CGCAGCGGATCATATAGCGCGGAGGCGGAGCCCGGAGAGCATCGAGCTGAGCGGC 945
QY 1204 GTGCTGACGTGCTGGCGGCTGTGTGGGACTGGGCAAGACCGTCAAGCGCTCGCCAAAC 1263
DB 946 ACCTTCTGTGGCGCCGAGGCCCCAGGTGAGCTGGGCGGCTTGGGCGGCGGCGCGCGAG 1005
QY 1264 GCGCCCAACATGAGGAGCAGCGCGCTCTGTGGGACAGCAACATGCTCATCTCACTTCTGTG 1323
DB 1006 GCGCGGAGCTGAGCGCCAGCGCCGCGGAGTCTCATGATCTCAAGGACCTCAAGCAG 1065
QY 1324 AAGACGGGCGCCAAAGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1383
DB 1066 AAGACCGCGCAAGGATCTGGAGCAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1125
QY 1384 AACAGGGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1443
DB 1126 CACGACGAAAGAGCGCGCTTCTACCGCATCAGGCGCACGC-----GGCTGTATGACC 1179
QY 1444 GCGGAGCGCATCCCACTTGAAGAACTACATCGCGCGCACCATGAGCGCGTGGCGGAGAAC 1503
DB 1180 GCGCGCGGAAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1239
QY 1504 TGCACTGTGGCAAGCAGAACTACTTCTACTACAACTGCTCAACCGGCAAGTTCCTGTGCG 1563
DB 1240 GAGGCGCGGCGGAGGAGCAGAGCAGCGCGCGCATCTTCTTTCGAGCGCTAGCCTC 1299
QY 1564 GACAACTGCCCCACTTACTCTGCGGAGGCGGCTTTCGCCACCTCAAGAGTGGCGTGTG 1623
DB 1300 TACCACTGTGTGAGAACTGCGGCTCGCTGTCTGTCTGCTGCTGCTGCTGCTGCTGCTG 1359
QY 1624 GACAACTGACCGCTTCCACCAACTTCTTCTCATGGAGA 1661
DB 1360 GGCACAGCACCTTCTACGTGAGCTACCGCACTGAGGA 1397
```

```
RESULT 10
US-11-183-624-3
; Sequence 3, Application US/11183624
; Publication No. US20050255518A1
; GENERAL INFORMATION:
; APPLICANT: Silos-Santiago, Immaculada
; TITLE OF INVENTION: Methods and compositions for the
; FILE OF INVENTION: treatment and diagnosis of pain disorders using 46556
; FILE REFERENCE: MPI01-272PIRM
; CURRENT APPLICATION NUMBER: US/11/183,624
; CURRENT FILING DATE: 2005-07-18
; PRIOR APPLICATION NUMBER: US/10/281,866
; PRIOR FILING DATE: 2002-10-28
; PRIOR APPLICATION NUMBER: 60/335,078
; PRIOR FILING DATE: 2001-10-31
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 4282
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (94)...(2859)
```

US-11-183-624-3

Query Match 2.8%; Score 56; DB 12; Length 4282;  
Best Local Similarity 44.2%; Pred. No. 0.00062;  
Matches 282; Conservative 0; Mismatches 350; Indels 6; Gaps 1;  
Qy 1024 CTGAGTTTTCAGGACGTGTGGCAGCTGTTCGGCGAGGGCGTGACCCCGCGCATTTGAGGAG 1083  
Db GAGGTGTGGAGGGCTGTGTGACCCCTGGTCTTCTTCCGGTGTGGGTATTTCGCTGG 825  
Qy 1084 CTGTACAGAGAGTGTGGCGCCCTTCTGTGCAAAACAGCAGCAAACTTCTGTGTCGAAG 1143  
Db ATGGCCGACAGCGGCTGCTCTTCTACAGTACGTGTACAGGCTTACCGCAGCCGACCA 885  
Qy 1144 CGCTCTGTGTACTTCCAGCAGCGCTGTATACAGGGCGGCATGGCAAGCTGTGTGG 1203  
Db CGCAGCGGCATCATATAGCGCGGAGGGGACCCCGGAAGAGCATCGAGCTGTGAGCGC 945  
Qy 1204 GTGTGAGTGCCTTGGCGGTGTGTGGGACTGTGGGCAAGACCGTCAAGCGCCTCGCCAAAC 1263  
Db AGTTCGTGGCGCGGAGGCCCGAGGTGAGTGTGGCGGCTGTGGCGCCCGCGCGAG 1005  
Qy 1264 GCGCCCAATATGAGGAGAGCGCGCTGTGTGGGACAGCAACATGTCTATCTTCACTTCGTG 1323  
Db GCGCGGAGTGTGAGCGCAGCGCGCGGAGGTATCCAGATCTCTCAAGGACCTCAAGCAG 1065  
Qy 1324 AAGNACGGGCCCGAGCGCTGT 1383  
Db AAGCACCAGGACAGGATCTGAGGAGCTGTGTGGGCAATCGCCAACTACTACGGGCTGTG 1125  
Qy 1384 AACAAAGCGGT 1443  
Db CACGAGAGAGAGCGCGGCTTCTACCGCATCAGGCGACGC-----GGCTGTATGCC 1179  
Qy 1444 GCGGACGGCATTCCTCATTTGAGAACTATCATCGCGGACCATATGAGCGGCGTGGCGGAAC 1503  
Db GCGCGCGGAAACGT 1239  
Qy 1504 TCGCAGTGTGGCAGCAGAACTACTTCTATCAACTGCTTCACTCGGCGCATTCCTCGCGC 1563  
Db GAGGCGCGGCGGAGGACGAGACGCGCGGACGCTTCCGCGCAGGCGGCGCGCGCGC 1299  
Qy 1564 GACAACTGTCCCACTTACCTGCGGAGCGGCGCTTCCGCGACCTTCAAGAGTGGCGTGTG 1623  
Db TACCAGTGTCTGGAGAACTGCGGTCTCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1300  
Qy 1624 GACAACTGACCGTCTCAACCAACTTCTTATGAGGA 1661  
Db GGCACAGCAGCCTTCTACGTGAGTACCGCACTGAGGA 1397

RESULT 11

US-11-096-568A-25531  
; Sequence 25531, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096.568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 25531  
; LENGTH: 1243  
; TYPE: DNA  
; ORGANISM: Zea mays subsp. mays  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (1)..(1243)  
; OTHER INFORMATION: Ceres Seq. ID no. 13491142  
US-11-096-568A-25531

Query Match 2.8%; Score 55.2; DB 9; Length 1243;  
Best Local Similarity 44.4%; Pred. No. 0.0009;  
Matches 271; Conservative 0; Mismatches 333; Indels 6; Gaps 1;  
Qy 1146 CTTCTGTGTACTTCAGCAGCGGCTGTACTACAGGGGGGATGGGCAAGCTGTGCTGGGT 1205  
Db CATGTGCGCCCTCTCACACCGGCTGGGACGGCAATGCGCGCCACAGCGTGCACACCGAT 255  
Qy 1206 GCTGTGAGTGTCTGGCGTGTGTGGACTGTGGCAAGACCGTCAAGCGGCTCGCCAAACGC 1265  
Db GCCCAAGAGGCTGAGCGGAGGTGCCATGTGACGGGCGGCGCGCGGGATCGGCGA 315  
Qy 1266 GCCCAACAATGAGGAGCAGCGCGTCTGTGGGACAGCAACATGCTCATCCACTTCGTGAA 1325  
Db GGCCATCGTGTGGCGTGTTCGCCAAGCAGCGGGCCCGGGTGTGTATCGCGACATCGACGA 375  
Qy 1326 GAACGGGCCCAAGCGGCTGTGTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1385  
Db CGCGCGGGGAGCGGCTGTGGCGTGTGGCGCTGTGGCGCGGCGGAGGTGAGCTTGTGTGGCTCGCA 435  
Qy 1386 CAAGGCGGCTGT 1445  
Db CGT 489  
Qy 1446 GAGCGGCGATCCCATTTAGAACTATCATCGCGCGCACCATGAGCGGCGTGTGGGAGAACTC 1505  
Db CGCGCGCGGCTCGACGTCTACTGCAACAAACGCGGGGTGTGTGGCGCGCAGACGCGCGC 549  
Qy 1506 GACGTGTGGCAGCAGAACTACTTCTATCAACTGCTTACCGGCGGCGGCGGCGGCGGCGG 1565  
Db CGCGCAGGAGCATCTGTCTTTCGACGCGCGCGGAGTTGACCGGCTGTGTGTGTGTGTGTGTGT 609  
Qy 1566 CAATGCGCCCACTTACCTGT 1625  
Db GCTGTGGCGCGCGGCTGT 669  
Qy 1626 CAACCTGACCGTGTCCCAACCAACTTCTTATGAGGAGCTCAAGCGCGGCGGCGGCGGCGG 1685  
Db GAGCATGTCTTCTGT 729  
Qy 1686 GTGTATTTGTATGACACAGT 1745  
Db CACCGCGCTCAAGCAGCGGCGGCTGT 789  
Qy 1746 GTGCTGTGGC 1755  
Db GCAAGGGGTC 799

RESULT 12

US-11-205-109-1  
; Sequence 1, Application US/11205109  
; Publication No. US20050287641A1  
; GENERAL INFORMATION:  
; APPLICANT: Farnet, Chris  
; APPLICANT: Zazopoulos, Emmanuel  
; APPLICANT: Staffa, Alfredo  
; TITLE OF INVENTION: GENE CLUSTER FOR RAMOPLANIN BIOSYNTHESIS  
; FILE REFERENCE: 3002-2US  
; CURRENT APPLICATION NUMBER: US/11/205.109  
; CURRENT FILING DATE: 2005-08-17  
; PRIOR APPLICATION NUMBER: US/09/976,059  
; PRIOR FILING DATE: 2001-10-15  
; PRIOR APPLICATION NUMBER: US 60/239,924  
; PRIOR FILING DATE: 2000-10-13  
; NUMBER OF SEQ ID NOS: 46  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1  
; LENGTH: 88421  
; TYPE: DNA  
; ORGANISM: Actinoplanes sp.  
; FEATURE:  
; NAME/KEY: misc.feature

/ LOCATION: (2077)..(3078)  
/ OTHER INFORMATION: ORF 1; positive strandedness  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (3118)..(4032)  
/ OTHER INFORMATION: ORF 2; positive strandedness  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (4038)..(5048)  
/ OTHER INFORMATION: ORF 3; positive strandedness  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (6665)..(5814)  
/ OTHER INFORMATION: ORF 4; negative strandedness  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (7703)..(6693)  
/ OTHER INFORMATION: ORF 5; negative strandedness  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (9464)..(8130)  
/ OTHER INFORMATION: ORF 6; negative strandedness  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (9691)..(10761)  
/ OTHER INFORMATION: ORF 7; positive strandedness  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (12751)..(10829)  
/ OTHER INFORMATION: ORF 8; negative strandedness  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (13617)..(12802)  
/ OTHER INFORMATION: ORF 9; negative strandedness  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (15203)..(13614)  
/ OTHER INFORMATION: ORF 10; negative strandedness  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (15591)..(15863)  
/ OTHER INFORMATION: ORF 11; positive strandedness  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (15880)..(19035)  
/ OTHER INFORMATION: ORF 12; positive strandedness  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (19032)..(39713)  
/ OTHER INFORMATION: ORF 13; positive strandedness  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (39713)..(65800)  
/ OTHER INFORMATION: ORF 14; positive strandedness  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (65826)..(66530)  
/ OTHER INFORMATION: ORF 15; positive strandedness  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (66546)..(67370)  
/ OTHER INFORMATION: ORF 16; positive strandedness  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (67384)..(70059)  
/ OTHER INFORMATION: ORF 17; positive strandedness  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (70099)..(70662)  
/ OTHER INFORMATION: ORF 18; positive strandedness  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (70659)..(71906)

/ OTHER INFORMATION: ORF 19; positive strandedness  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (73439)..(71964)  
/ OTHER INFORMATION: ORF 20; negative strandedness  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (74216)..(73563)  
/ OTHER INFORMATION: ORF 21; negative strandedness  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (75424)..(74213)  
/ OTHER INFORMATION: ORF 22; negative strandedness  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (75535)..(76464)  
/ OTHER INFORMATION: ORF 23; positive strandedness  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (78110)..(76449)  
/ OTHER INFORMATION: ORF 24; negative strandedness  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (79864)..(78107)  
/ OTHER INFORMATION: ORF 25; negative strandedness  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (81624)..(79861)  
/ OTHER INFORMATION: ORF 26; negative strandedness  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (81909)..(81682)  
/ OTHER INFORMATION: ORF 27; negative strandedness  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (82346)..(82062)  
/ OTHER INFORMATION: ORF 28; negative strandedness  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (82587)..(84446)  
/ OTHER INFORMATION: ORF 29; positive strandedness  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (84481)..(85548)  
/ OTHER INFORMATION: ORF 30; positive strandedness  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (85556)..(86845)  
/ OTHER INFORMATION: ORF 31; positive strandedness  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (87372)..(86803)  
/ OTHER INFORMATION: ORF 32; negative strandedness  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (87494)..(88420)  
/ OTHER INFORMATION: ORF 33; positive strandedness; N-terminus only  
US-11-205-109-1

Query Match 2.8%; Score 55.2; DB 12; Length 88421;  
Best Local Similarity 45.7%; Pred. No. 0.0011;  
Matches 275; Conservative 0; Mismatches 318; Indels 9; Gaps 2;  
QY 1334 CCAAGCCGCTGGTGTGCTCAAGTTCGTGAGCCCTGCTTCAACAAAGCCG 1393  
Db 60768 CCGGACGCTCGCCGACCTGCTCAACTTCGAGCTCTCGGCTTCGAGATCTTCGGC 60827  
QY 1394 TCGTGTGTTCCGGCGCGCGTCCCGGCAAGCAGTACGCTGTGATCAAGCGGACGGCA 1453  
Db 60828 CGCTGTGTCCGGCGCGCAGCATCGAGATCGTCACCGACCTGCTCGCCCTGGCCGACCCG 60887  
QY 1454 TCCCATTCAGAACTACATCGCGCGCACCATGAGCGGCTGGCGGAGAACTCCACGTGC 1513



```
Db 60888 CCTCCCGCGCTGGAGGCGCAGCTGTCAGCGCGGTGCGTGGCGGTTCTCGCGGTC 60947
Qy 1514 GCAAGCGAAGTACTTCTTACTACACTGCTCTACCGGCAAGTTCCTGCGGACACTGCC 1573
Db 60948 TCAGCCGGGCGCATCGCGCGCGCACCCGCGAGCGTGGTGGCGCGGCGCGCTGA 61007
Qy 1574 CCACCTACCTGCGGAGGCGGCTTTCGCGGCGGCTCAAGAGTGGCGTGGGACAACTGA 1633
Db 61008 CCGCGGAGTGGTGAACGCCACCGTGCCTGCGCGCTGCGCGGTGCGGAGTCCAACTCT 61067
Qy 1634 CCGTCTCCAACTTCTTCATGAGGAGTCAAGCGCGCACTTACACCAAGGTGATTC 1693
Db 61068 ACGGCGGACCGAGGCGACGCTCTACTCGACCGCTTGGCACACCGACCGGAGCTGACCG 61127
Qy 1694 TGATGACCAAGTGGACTGCTGATATGCGCGTGGGCAACG---AGCTGGCGGAGTGCC 1750
Db 61128 GCGGCGCGCGCGCATCGGCGCGCGCTCAACCAACCCGCGCTAGTCTCTCGAGACC 61187
Qy 1751 TGCCCAAGCAGGTTGGCGCGGCGGCATCGTCTGCGGCTCGCGCTCTCAGCCCGC 1810
Db 61188 GTCTACCGCGTGGCGCGCGCGTGGTGGCGAGCTCTACTGCGCGCGCGAGCTGG 61247
Qy 1811 CCTACGCGGAGCTGATCCAGAAGCGCGGCTTCGACGTCGCTGCATCCGCGCGCACTC 1870
Db 61248 CCGCGGCTATCTGGGCGCGCGCGCTGACCGCGGAGCGCTTGGT-----GGCCTGCC 61301
Qy 1871 AGGCTACATGAGCGCGTCAACATGTACAGCTTCTTACATGCGCGCGGAGGCGG 1930
Db 61302 CGTTGCGCGCGCGGCGGAGCATGTACCGCACCGCGCGCGGTCGCGTGGAAACCG 61361
Qy 1931 CC 1932
Db 61362 AC 61363

RESULT 13
US-11-205-109-1/c
; Sequence 1, Application US/11205109
; Publication No. US20050287641A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Zazopoulos, Emmanuel
; APPLICANT: Staffa, Alfredo
; TITLE OF INVENTION: GENE CLUSTER FOR RAMOPLANIN BIOSYNTHESIS
; FILE REFERENCE: 3002-2US
; CURRENT APPLICATION NUMBER: US/11/205,109
; CURRENT FILING DATE: 2005-08-17
; PRIOR APPLICATION NUMBER: US/09/976,059
; PRIOR FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: US 60/239,924
; PRIOR FILING DATE: 2000-10-13
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 88421
; TYPE: DNA
; ORGANISM: Actinoplanes sp.
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2077)..(3078)
; OTHER INFORMATION: ORF 1; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (3118)..(4032)
; OTHER INFORMATION: ORF 2; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4038)..(5048)
; OTHER INFORMATION: ORF 3; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (6665)..(5814)
; OTHER INFORMATION: ORF 4; negative strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (7703)..(6693)
; OTHER INFORMATION: ORF 5; negative strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (9464)..(8130)
; OTHER INFORMATION: ORF 6; negative strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (9691)..(10761)
; OTHER INFORMATION: ORF 7; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (12751)..(10829)
; OTHER INFORMATION: ORF 8; negative strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (13617)..(12802)
; OTHER INFORMATION: ORF 9; negative strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (15203)..(13614)
; OTHER INFORMATION: ORF 10; negative strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (15591)..(15863)
; OTHER INFORMATION: ORF 11; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (15880)..(19035)
; OTHER INFORMATION: ORF 12; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (19032)..(39713)
; OTHER INFORMATION: ORF 13; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (39713)..(65800)
; OTHER INFORMATION: ORF 14; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (65826)..(66530)
; OTHER INFORMATION: ORF 15; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (66546)..(67370)
; OTHER INFORMATION: ORF 16; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (67384)..(70059)
; OTHER INFORMATION: ORF 17; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (70059)..(70662)
; OTHER INFORMATION: ORF 18; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (70659)..(71906)
; OTHER INFORMATION: ORF 19; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (73439)..(71964)
; OTHER INFORMATION: ORF 20; negative strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (74216)..(73563)
; OTHER INFORMATION: ORF 21; negative strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (75424)..(74213)
; OTHER INFORMATION: ORF 22; negative strandedness
; FEATURE:
```

```

; NAME/KEY: misc feature
; LOCATION: (75535)..(76464)
; OTHER INFORMATION: ORF 23; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (78110)..(76449)
; OTHER INFORMATION: ORF 24; negative strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (79864)..(78107)
; OTHER INFORMATION: ORF 25; negative strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (81624)..(79861)
; OTHER INFORMATION: ORF 26; negative strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (81909)..(81682)
; OTHER INFORMATION: ORF 27; negative strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (82346)..(82062)
; OTHER INFORMATION: ORF 28; negative strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (82587)..(84446)
; OTHER INFORMATION: ORF 29; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (84481)..(85548)
; OTHER INFORMATION: ORF 30; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (85556)..(86845)
; OTHER INFORMATION: ORF 31; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (87372)..(86803)
; OTHER INFORMATION: ORF 32; negative strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (87494)..(88420)
; OTHER INFORMATION: ORF 33; positive strandedness; N-terminus only
; OTHER INFORMATION: ORF 33; positive strandedness; N-terminus only
; US-11-205-109-1

Query Match      2.8%; Score 54.8; DB 12; Length 88421;
Best Local Similarity 41.2%; Pred. No. 0.0014;
Matches 383; Conservative 0; Mismatches 547; Indels 0; Gaps 0;

QY 632 TCGGCCCCGAGCGCGCTACTTGGAGCAGAGAGCTGGAGCGCGCTGTGGAGCAGAAC 691
DB 77362 TGGGCGCGTGGACAGCGCGTGGCGTGGCCACCGAGATGGCGATGGAGCGCGCTGT 77303
QY 692 CCAGGGTTGATCCCTAGTCCGCTGGCTGGCGCGCCCTACTAGTGTGATTTGGCC 751
DB 77302 ACGGGCGCGCGTGGCGGACCTCGCTGACGTGGCGGACCAACATCGCGCGCGCTATGCTG 77243
QY 752 GCCTGCCAGCGTTGGCGACGCGCTTGACAGGAGCGCGTGGAGCGCGCGCCCATGTTCC 811
DB 77242 CCTGCTACCGTTCAGCTTCTTCGGGCGTGGCGCTGGCGCGCTGACCTCTGCGCG 77183
QY 812 CGCCACCTTCTCTACACGAGTCTGGTGGAGGAGCCCGAGCGCGGATATGAGGTGATGG 871
DB 77182 AGGCCAGCGCGGTTCAGCGCGCGCGTGAAGTACCTGACCGCGCAGATGCTGTGACG 77123
QY 872 AGATCAACCCCAAGACAGCGTGTGACCTCTGACTAGCGCGCGCTGCAATGCCCTGAAC 931
DB 77122 CGATCAACGATCTCGCTGCTGGCGCGCGAGGTTATCTGCGCGAGGCGCGGTACG 77063
QY 932 TGCTGTGAGGGGCGCGCGCGAGTGTGCTGGTGGAGTCTCAACCGCGCGAGTTCGCGCG 991
DB 77062 CGATGTTCCAGAGTGGTTCGCGACGCGCGCGCGCTCTTTCGCGACGCTGTCCGCGG 77003

992 TTCTGAGCTGAAGAGGTGGCCATTTAGCAGCTGGAGTTTGAAGAGCTGTGGAGCTGT 1051
77002 CGGCTTGCCTGTGTGATGCTGCTGCCGACACCTGCCCGGTGGCGCGCTGTGGACGG 76943
QY 1052 TGGCGAGGGCGTGCACCCCGCATTTAGGAGCTGTACGAGAGAGCTGGCGCCCTTCC 1111
DB 76942 CCAGAGAGCGCGCGCGGACACGTTTCACTTCGCGGCGAGCTGTCCCGCTTGACT 76883
QY 1112 TGTTCGCAACACGACCACTTCTGCTGTCGAAGCGCTCTGTGTTACTTCCAGCAGCGCTGT 1171
DB 76882 TCTCCCGGCTGTGTCAGCGGCGATCGTGGCGACCCCGCTGGCGCGCTCTTCCAGACTCGT 76823
QY 1172 ACTACAGGGCGCGCATGGGCAAGCTGTGTGGTGTGTGAGTGTCTGCGCGCTGTGTGCTGG 1231
DB 76822 GGCACGACGAGGGCGCGCTGCGCGGTTTCCCGAGCGGTTCCACCGGAGCTGACCGGCC 76763
QY 1232 GACTGGGCAAGACCGTCAAGCGCTCGCCAAAGCGCCCAATGGAAGGAGCAGCGCGTC 1291
DB 76762 TGGGAGACGCTGCGCGGAGCTGGGCGCGCGCCACATCATGATCGACGCCACCGCGCG 76703
QY 1292 TGTGGACAGCAACATGCTCATCTTCTGTTGAAGAACGGGCGCCCAAGCGCTGTGTGGC 1351
DB 76702 CCTTCGCGCTGGCGCGACCGGTACACCGTGTCTGTGGCGCGCTTGGCGCTGTGGGTGT 76643
QY 1352 TGTTCGTTCAAGTTCGTGAGCTGTGTCTCTTCAAGGCGCGTGTGTGTGGTGTGGCGCG 1411
DB 76642 GCGCGCGCGTGTGGCGCGCTTCCACCGCGCGCTGTGTGGCGCTTCTGGAGCGGCTGGCG 76581
QY 1412 GCGTGGCGGCAAGCAGTACGCGCTCATCAAGCGGACGCGCATCCCATTTGAGAACTACA 1471
DB 76582 GCGGCTCGCGCGGAGCGCGTGTGAGCTGGCGCGCGCTGCGCGCGCTGCTGGACCTGAGCGCGCTC 76523
QY 1472 TCGCGCGCACCATGAGCGCGTGTGGCGGAGAACTCGCACGCTGGCGCAAGCAACTACTTCT 1531
DB 76522 TCTTCGAGATGGCGCGACCGGTTACGACCGCGCTGTGTGGACCTGAGCGCGCGCTC 76463
QY 1532 ACTACAACTGCTTACCGCGCAAGTTTCTGTC 1561
DB 76462 AGTCCCGCGCTGACCGCGCGGATCAGGC 76433

RESULT 14
US-11-096-568A-23551
; Sequence 23551, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; TITLE OF INVENTION: Thery
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 23551
; LENGTH: 1283
; TYPE: DNA
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(1283)
; OTHER INFORMATION: Ceres Seq. ID no. 12413628
; US-11-096-568A-23551

Query Match      2.8%; Score 54.6; DB 9; Length 1283;
Best Local Similarity 47.3%; Pred. No. 0.0012;
Matches 185; Conservative 0; Mismatches 184; Indels 0; Gaps 0;

QY 1581 CTGCGCGAGGGCGCGCTTCCCAACCTCAAGAGTGGCGTGGTGGAGCAACCTGACCGTCTC 1640
DB 340 CGTGGCGGTGTGGGATCGCTTGGCGCGCTTGGCGCGATGATCTTCTGCTCTGCTACTG 399
QY 1641 CACCAACTTCTTCATGAGAGCTCAAGCGCGCACCTACACCAAGTGTTCATGATGA 1700
```

Db 400 CACGCGCGCATCTCGGGGGGCACATCAACCGCGCTCACCTTCGGCCTCTTCTTGGC 459  
Qy 1701 CCAGTGGATGGCTGATGATGCGGTGGCAAGAGCTGGCGAGTGGCTTGGCCAGCA 1760  
Db 460 GCGCAAGGTCTCCCTGGTGGCGGCTCTTACATCTGTCGCGAGTGGCTTCGGCGCCAT 519  
Qy 1761 GGTTCGCGCGGGGATCTGTCATCTGGCGCTCGGCTCTCAGCGCGCTTACGCCGA 1820  
Db 520 CTGCGGGCTCGGCTCGTCAAGGGTTCAGAGGGCTCTTTCGACAGGTACGGCGGG 579  
Qy 1821 GCTGATCAGAAGCGGGCTTCAGCGTGGCTGATCCGCGGCCACTCAGGGCTACAT 1880  
Db 580 CGCAACTGCTCGCTCCGGCTACTCCGCGGCACCGGCTCGGCGCGAGATCATCGG 639  
Qy 1881 GGACCGGCTCAAGTGTACAGTCTCTTCTACATGGCGCGCGGAGGCG 1929  
Db 640 CACCTTCGTGCTGCTTACACCGCTCTTCTCGCCACCGACCCCAAGCGC 688

## RESULT 15

US-11-167-048-2  
; Sequence 2, Application US/11167048  
; Publication No. US20050266487A1  
; GENERAL INFORMATION:  
; APPLICANT: Batany, Francis  
; APPLICANT: Cao, Weiguo  
; APPLICANT: Tong, Jie  
; TITLE OF INVENTION: HIGH FIDELITY THERMOSTABLE LIGASE AND USES THEREOF  
; FILE REFERENCE: 19603/2618  
; CURRENT APPLICATION NUMBER: US/11/167,048  
; CURRENT FILING DATE: 2005-06-24  
; PRIOR FILING DATE: 1998-10-30  
; PRIOR APPLICATION NUMBER: 60/106,461  
; PRIOR FILING DATE: 1999-10-29  
; PRIOR APPLICATION NUMBER: PCT/US99/25437  
; PRIOR FILING DATE: 1999-10-29  
; PRIOR APPLICATION NUMBER: 09/830,502  
; PRIOR FILING DATE: 1999-10-29  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 2025  
; TYPE: DNA  
; ORGANISM: Thermus sp.  
US-11-167-048-2

Query Match 2.8%; Score 54; DB 12; Length 2025;  
Best Local Similarity 48.9%; Pred. No. 0.0018;  
Matches 175; Conservative 0; Mismatches 180; Indels 3; Gaps 1;  
Qy 1468 TACATCGCGCACCATGGAGCGGTGGCGGAGAACTCGCACGTGGCGCAAGCAGAACTAC 1527  
Db 907 TACACCGCGCACCCCGCGCTTCGCCCTCGCTTACAAGTTCCGCGCGCATCACCCCGTGGC 966  
Qy 1528 TTCTACTACAACTGCCTCACCGCAAGTTCCTGCGCGACAACTGCCCCACCTACTCTGCGC 1587  
Db 967 ACCGCGCTCTCTCCGTGGCTTCCAGGTGGGGCGGACCGGGCGCATCACCCCGTGGC 1026  
Qy 1588 GAGCGCGCTTCGCCACCTCAAGAGTGGGTGTGAGCAACCTGACCGCTTCACCAAC 1647  
Db 1027 GTTCTGGAGCCGCTTTCATAGAGGCGAGGTGAGCGGGTCACTCCCTCCACACAG 1086  
Qy 1648 TTCTTCATGGAGGAGCTCAAGCGCGCATCTACACCAAGGTGATTCTGTATGGACAGTGT 1707  
Db 1087 AGCTTCATTGAGGAGCTGGAGCTGGCATCGGCACTGGGTG---CTGGTCCACAGCG 1143  
Qy 1708 GACTGGCTGATATGCCCGTGGCCAAAGAGCTGGCGGCTGCTGGCCAAAGAGGTTGCG 1767  
Db 1144 GCGGGGTGATTCCGAGGTGCTGAGGTCTGAAAGAGCGCCGACCGGGGAGAGAAG 1203  
Qy 1768 CCGGGCGGATCGTGTATCTGGCGCTTCGCGCTCCCTCAGCCGCGCTTAGCGCGAGCTGA 1825  
Db 1204 CCCATCATCTGGCCCGAGAACTGCCCCGAGTGGCGGCCACCGCCTCATCAAGAGGGGA 1261

Search completed: March 14, 2006, 01:29:35  
Job time : 661 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: March 14, 2006, 01:18:41 ; Search time 34.5 Seconds  
(without alignments)  
4959.249 Million cell updates/sec

Title: US-10-620-914-44

Perfect score: 3619

Sequence: 1 aggggtcggctcgtagcgg.....gagccagaagagacaactaa 1947

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 4886326

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=x1h  
-O=/abss/ABSSWEB.spool/US10620914/runat\_13032006\_101923\_5407/app\_query.fasta 1  
-DB-A Genesecp -QWTF=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0 -LOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blomum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs05h  
-USER=US10620914 @CGN 1.1 157 @runat\_13032006\_101923\_5407 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A Genesecp 21.\*

1: Genesecp1980s.\*  
2: Genesecp1990s.\*  
3: Genesecp2000s.\*  
4: Genesecp2001s.\*  
5: Genesecp2002s.\*  
6: Genesecp2003as.\*  
7: Genesecp2003bs.\*  
8: Genesecp2004s.\*  
9: Genesecp2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3463	95.7	648	9	ADW72748 Chlamydomonas reinhardtii
2	891	24.6	752	8	ADP98846 C. albica
3	850.5	23.5	908	9	ADW72753 Neurospora
4	842.5	23.3	845	8	ADR86056 Aspergillus
5	462	12.8	19938	6	ADW72746 Chlamydomonas reinhardtii
6	453	12.5	19938	6	ADW72747 Chlamydomonas reinhardtii
7	440.5	12.2	19938	6	ADW72748 Chlamydomonas reinhardtii
8	439	12.1	19938	6	ADW72749 Chlamydomonas reinhardtii
9	437	12.1	19938	6	ADW72750 Chlamydomonas reinhardtii

10	435.5	12.0	19938	6	AB988398 Streptomyces
11	416.5	11.5	19938	6	ABP76678 Streptomyces
12	407.5	11.3	19938	6	ABP76682 Streptomyces
13	398.5	11.0	19938	6	ABP76678 Streptomyces
14	386.5	10.7	1212	2	AAW87503 Human N-m
15	386	10.7	1212	2	AAW87503 Human N-m
16	386	10.7	1212	2	AAW87503 Human N-m
17	386	10.7	1212	2	AAW87503 Human N-m
18	386	10.7	1212	2	AAW87503 Human N-m
19	385	10.6	1061	2	AAW87504 Human N-m
20	384.5	10.6	1081	3	AAW87504 Human N-m
21	384.5	10.6	1081	3	AAW87504 Human N-m
22	384.5	10.6	1081	3	AAW87504 Human N-m
23	384.5	10.6	1081	3	AAW87504 Human N-m
24	381.5	10.5	416	9	ADW72756 Rhodospirillum rubrum
25	379.5	10.5	416	9	ADW72706 Rhodospirillum rubrum
26	379.5	10.5	416	9	ADW72706 Rhodospirillum rubrum
27	378.5	10.5	1615	4	AAW87503 Human N-m
28	374.5	10.3	663	7	ABO82097 Pseudomonas
29	373	10.3	19938	6	AB988398 Streptomyces
30	372.5	10.3	601	7	ABM86504 Rice abio
31	370	10.2	900	4	ABG03533 Novel hum
32	361.5	10.0	1017	4	AAW87503 Human N-m
33	356	9.8	440	6	ADA48432 Rice abio
34	352.5	9.7	601	7	ABM86504 Rice abio
35	352	9.7	599	4	ABG03530 Novel hum
36	346.5	9.6	1017	4	AAW87503 Human N-m
37	346.5	9.6	1615	4	AAW87503 Human N-m
38	333	9.2	1592	4	AAW87503 Human N-m
39	332.5	9.2	863	7	ABO82097 Pseudomonas
40	332.5	9.2	19938	6	ABP76681 Streptomyces
41	332	9.2	536	7	ABO82097 Pseudomonas
42	330.5	9.1	1592	4	AAW87503 Human N-m
43	329.5	9.1	412	7	ABM87971 Rice abio
44	329.5	9.1	412	7	ABM87926 Rice abio
45	329.5	9.1	1212	2	AAW87503 Human N-m

#### ALIGNMENTS

#### RESULT 1

ADW72748

ID ADW72748 standard; protein; 648 AA.

XX

AC ADW72748;

XX

DT 21-APR-2005 (first entry)

XX

DE Chlamydomonas reinhardtii Btal gene, protein.

XX

KW Phospholipid synthesis; transgenic plant; fertilizer; enzyme;

XX

KW betaine lipid.

XX

OS Chlamydomonas reinhardtii.

XX

PN WO2005009115-A2.

XX

PD 03-FEB-2005.

XX

PF 15-JUL-2004; 2004WO-US022789.

XX

PR 16-JUL-2003; 2003US-00620914.

XX

PA (UNNS ) UNIV MICHIGAN STATE.

XX

PI Benning C, Riekhof W;

XX

DR WPI; 2005-112975/12.

XX

DR N-PSDB; ADW72746, ADW72747.

XX

PT New composition comprising a purified DNA having an oligonucleotide sequence encoding a protein, useful in producing betaine lipid compounds e.g., Diacylglycerol-O-4'-(N,N,N-trimethyl) homoserine (DGTS).



Db 561 ThrLysValIleLeuMetAspHisValAspTrpLeuAspMetProValAlaAsnGluLeu 580  
 QY 1741 GCCGAGTCTGCGCAAGCAGGTTGCGCGGGCGCATCTCATCTGGCGCTCCGCTCC 1800  
 Db 581 AlaGluCysLeuAlaLysGlnValAlaProGlyGlyIleValIleTrpArgSerAlaSer 600  
 QY 1801 CTCAGCCCGCCCTACCGCGAGCTGATCCAGAGCGCGGCTTCGACGCTGCATCCGC 1860  
 Db 601 LeuSerProProTyrAlaGluLeuIleGlnLysAlaGlyPheAspValArgCysIleArg 620  
 QY 1861 CGCGCCACTCAGGCTACATGACCGCGTCAACATGTACAGCTCCCTTACATGGCGCCG 1920  
 Db 621 ArgAlaThrGlnGlyTyrMetAspArgValAsnMetTyrSerSerPheTyrMetAlaArg 640  
 QY 1921 CGCAAGCGCGCCCAAGAGGACAAC 1944  
 Db 641 ArgLysGlyAlaLysLysAspAsn 648

RESULT 2  
 ADP98846  
 ID ADP98846 standard; protein; 752 AA.  
 XX  
 AC ADP98846;  
 XX  
 DT 23-SEP-2004 (first entry)  
 XX  
 DE C. albicans specific gene, orf6.3438, protein sequence.  
 XX  
 KW Diploid fungal cell; allele; gene disruption cassette;  
 KW promoter replacement fragment; antifungal; fungicide; gene therapy;  
 KW infection; Candida albicans.  
 XX  
 OS Candida albicans.  
 XX  
 PN WO2004056965-A2.  
 XX  
 PD 08-JUL-2004.  
 XX  
 PF 19-DEC-2003; 2003WO-US040618.  
 XX  
 PR 19-DEC-2002; 2002US-0434832P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 PA (ELIT-) ELITRA CANADA LTD.  
 XX  
 PI Roemer T, Jiang B, Boone C, Bussey H;  
 XX  
 DR WPI; 2004-500296/47.  
 DR N-PSDB; ADP98536.  
 XX  
 XX Constructing a strain of diploid fungal cells in which both alleles of a  
 PT gene are modified comprises modifying the alleles of a gene in the fungal  
 PT cells by recombination using a gene disruption cassette and a promoter  
 PT replacement fragment.  
 XX  
 PS Claim 44; SEQ ID NO 7021; 163pp; English.  
 XX  
 XX The invention relates to a novel method for constructing a strain of  
 CC diploid fungal cells in which both alleles of a gene are modified. The  
 CC method comprises modifying the alleles of a gene in diploid fungal cells  
 CC by recombination using a gene disruption cassette and a promoter  
 CC replacement fragment. The invention further comprises: assembling a  
 CC collection of diploid fungal cells each of which comprises modified  
 CC alleles of a different gene; a strain of diploid fungal cells comprising  
 CC modified alleles of a gene, where the first allele of the gene is  
 CC inactivated by a gene disruption cassette comprising a nucleotide  
 CC sequence encoding an expressible selectable marker; and the expression of  
 CC the second allele of the gene is regulated by a heterologous promoter  
 CC that is operably linked to the coding region of the second allele of the  
 CC gene, and where the gene encodes the polypeptide mentioned above; a  
 CC collection of diploid fungal strains comprising the diploid strains cited  
 CC above, where substantially all the different genes that encode the above

CC amino acid sequences are modified and are present in different diploid  
 CC strains in the collection; a nucleic acid molecule microarray comprising  
 CC nucleic acid molecules, where each nucleic acid molecule comprises a  
 CC nucleotide sequence that is hybridizable to a target nucleotide sequence  
 CC comprising any of the 310 nucleotide sequences listed in the  
 CC specification (ADP98516-ADP98825); identifying a gene that is essential  
 CC to the survival or growth of a fungus, that contributes to the virulence  
 CC and/or pathogenicity of a fungus, or that contributes to the resistance  
 CC of a diploid fungus to an antifungal agent; identifying an antifungal  
 CC agent that inhibits the growth of a diploid fungus, or a therapeutic  
 CC agent for treatment of a mammalian disease; correlating changes in the  
 CC levels of proteins or gene transcripts with the inhibition of growth or  
 CC proliferation of a diploid fungal cell; a purified or isolated nucleic  
 CC acid molecule comprising a nucleotide sequence encoding a gene product  
 CC required for proliferation of Candida albicans, where the gene product  
 CC consists of any of the above-mentioned amino acid sequences; a vector  
 CC comprising a promoter operably linked to the nucleic acid molecule cited  
 CC above; a host cell containing the vector; a purified or isolated  
 CC polypeptide comprising any of the 61 amino acid sequences given in the  
 CC specification (ADP96718-ADP96778); a fusion protein comprising a fragment  
 CC of a first polypeptide fused to a second polypeptide, the fragment  
 CC consisting of at least 6 consecutive residues of any of ADP98826-ADP99135  
 CC ; producing a polypeptide; identifying a compound which modulates the  
 CC activity of a gene product encoded by a nucleic acid comprising any of  
 CC ADP98516-ADP98825; eliciting an immune response in an animal; a strain of  
 CC Candida albicans, where a first allele of a gene comprising any of  
 CC ADP98516-ADP98825 is inactive and a second allele of the gene is under  
 CC the control of a heterologous promoter; identifying a compound or binding  
 CC partner that binds to the polypeptide comprising any of ADP98826-  
 CC ADP99135, or its fragment; identifying a compound having the ability to  
 CC inhibit growth or proliferation of Candida albicans; inhibiting growth or  
 CC proliferation of Candida albicans cells; manufacturing an antimycotic  
 CC compound; treating an infection of a subject by Candida albicans;  
 CC preventing or containing contamination of an object by Candida albicans,  
 CC or for preventing or inhibiting formation on a surface of a biofilm  
 CC comprising Candida albicans; a pharmaceutical composition comprising a  
 CC therapeutic amount of an agent which reduces the activity or level of a  
 CC gene product encoded by a nucleic acid comprising any of ADP98516-  
 CC ADP98825 in a pharmaceutical carrier; an antibody preparation which binds  
 CC the polypeptide; methods for evaluating a compound against a target gene  
 CC product encoded by any of ADP98516-ADP98825; identifying an antimycotic  
 CC compound; a computer or a computer readable medium that comprises at  
 CC least one of the nucleotide sequences mentioned in the specification or  
 CC at least one amino acid sequence selected from ADP98826-ADP99135; a  
 CC method assisted by a computer for identifying a putatively essential gene  
 CC of a fungus; and a protein array comprising proteins, where at least one  
 CC protein comprises an amino acid sequence or a portion of an amino acid  
 CC sequence selected from ADP98516-ADP98825. The novel methods and  
 CC compositions have fungicide activity. The compositions may be used in  
 CC gene therapy. The composition and methods are useful for drug screening  
 CC purposes or for diagnosing, preventing or treating infections associated  
 CC with Candida albicans. These may also be used for constructing strains  
 CC useful for identification and validation of gene products as effective  
 CC targets for therapeutic intervention, for identifying and validating gene  
 CC products as effective targets for therapeutic intervention, and for  
 CC collecting identified essential genes. This sequence represents the  
 CC protein of a Candida albicans fungal specific gene of the invention.  
 CC NOTE: This sequence was downloaded from an electronic sequence listing  
 CC provided on the WIPO website.

XX SQ Sequence 752 AA;

Alignment Scores:

Pred. No.:	1,4e-57	Length:	752
Score:	891.00	Matches:	229
Percent Similarity:	46.6%	Conservative:	98
Best Local Similarity:	32.7%	Mismatches:	232
Query Match:	24.6%	Indels:	142
DB:	8	Gaps:	21

US-10-620-914-44 (1-1947) x ADP98846 (1-752)

QY 151 CTGAGAGCTTCTACGGCGCCAGCGCGTTCCTTT----- 186







QY 1123 AGCCACAACTTCTGGTCCAAAGCGCTCTGTTCTCCAG-----CACGGCGCTG 1170  
 Db 598 AlaPheGlnTyrTrpLeuSerAsnAlaHisIlePheThrAspProAlaGlyArgGlyLeu 617  
 QY 1171 TACTACACAGCGCGCATGGGCAAGCTGTGTGGTGTCTGCAGCTGCTGGCGGTGGTCTG 1230  
 Db 618 TyrAspThrGlyGlySerArgTyrAlaIleArgPheArgTrpIleSerThrLeuPhe 637  
 QY 1231 GGACTGGCAAGACCGTCAAGCGCTGCCCAACGCGCCCAATGGAGGAGCGCCGT 1290  
 Db 638 PheCysArgSerAlaValArgLeuSerThrProThrLeuGluGlyGlnArgSer 657  
 QY 1291 CTGTGGGACAGCAACATCTCATCTTCTGTTGAAGACGGCGCCCAAGCGCTGGTGTGG 1350  
 Db 658 IleTyrHisThrIle-----ArgProCys----- 666  
 QY 1351 CTGTTCGTCAAGTTCGTGAGC---CTGGTCTCTTCAACAGCGCGTGTGTGGTGGC 1407  
 Db 667 LeuLeuAsnArgPheValAsnGlyLeuValLeuSerSerAspAlaPheLeuTrpSerAla 686  
 QY 1408 GCGCGGTGCGGCAAGCAGTACGCTGCTCATCAAGCGGAC----- 1449  
 Db 687 LeuGlyValProLysAsnGlnValAlaMetIleGluAlaAspTyrHisArgSerIle 706  
 QY 1450 -----GGCATCCCATTTGAGAC 1467  
 Db 707 SerSerSerThrThrProSerSerLysGluLysProSerArgAlaGluAlaIleLeuHis 726  
 QY 1468 TATATCGCGGCACCATGAGCGGTGGCGGAGAACTTCGCAGTGGCCAAAGCAAGACTAC 1527  
 Db 727 TyrThrThrSerThrLeuAspProValLeuSerThrSerHisLeuAlaSerAspAsnPro 746  
 QY 1528 TTCTACTACACTGCTCAGCGCAAGTTCCTGGCGCACAACTGCCCACTACCTACCTCGC 1587  
 Db 747 TyrTyrLeuValCysValLeuGlyGlnTyrThrArgGlnCysHisProAspTyrLeuSer 766  
 QY 1588 GAGCGCGCTTCGCCACCTCAAGAGT---GGCGTGTGGACAACTTCACCGCTCCACC 1644  
 Db 767 ProIleAlaHisSerIleLeuSerAlaProGlyAlaPheAspGlyLeuArgIleHisThr 786  
 QY 1645 AACTTCTTCATGGAGGAGCTC-----AAAGCGCGCACCTACACCAAGGTGATCTG 1695  
 Db 787 AspGluIleGlnGluValLeuAlaArgPheGlnProGlyThrLeuThrValAlaValVal 806  
 QY 1696 ATGACACACGTGAGCTGCTGATATGCC----- 1725  
 Db 807 MetAspSerMetAspTrpPheAspProProSerProGluGluLysGluGlyArgGly 826  
 QY 1726 GTGGCCAAACAGCTGGCGGAGTGTGCGCAAGCAGGTGTGGCGCGCGCATCGTCACTC 1785  
 Db 827 LysAlaArgGluGlnValArgArgLeuAsnArgAlaLeuLysValGlyGlyLysValLeu 846  
 QY 1786 TGGCGCTCCGCTCCCTCAGCCCTCAGCCGAGCTGTCCAGAGCGGGCTTCGAC 1845  
 Db 847 LeuArgSerAlaGlyValGluProTrpTyrValArgValPheValGluGluGlyPheGly 866  
 QY 1846 GTGGCGTGCATC-----CGCGCGCCACTCAGGCTACATGAGCGCGCTC 1890  
 Db 867 AlaArgValGlyCysArgGluSerGlyArgGlyAspGlnGluCysIleAspArgVal 886  
 QY 1891 AACATGTACAGCTCTTCTACATGGCGCGCGG 1923  
 Db 887 AsnMetTyrAlaSerCysTrpIleLeuGluLys 897

RESULT 4  
 ADR86056  
 ID ADR86056 standard; protein; 845 AA.  
 XX  
 AC ADR86056;  
 XX  
 DT 04-NOV-2004 (first entry)  
 XX

DE Aspergillus fumigatus essential gene protein #106.  
 XX Fungicide; Aspergillus fumigatus infection; Farmer's lung disease;  
 KW drug screening.  
 KW  
 XX Aspergillus fumigatus.  
 XX WO2004067709-A2.  
 XX 12-AUG-2004.  
 XX 16-JAN-2004; 2004WO-US001099.  
 XX 17-JAN-2003; 2003US-0441281P.  
 XX 13-JUN-2003; 2003US-0478196P.  
 XX (ELIT-) ELITRA PHARM INC.  
 XX (ELIT-) ELITRA CANADA LTD.  
 XX Jiang B, Hu W, Lemieux S, Roemer T;  
 XX WPI; 2004-594200/57.  
 XX N-PSDB; ADR85469.  
 XX New purified or isolated Aspergillus fumigatus nucleic acid molecule  
 XX encoding a gene product, useful for diagnosing and/or treating invasive  
 XX fungal infections, such as Farmer's lung disease.  
 XX Claim 1; SEQ ID NO 3106; 164pp; English.  
 XX The present invention relates to Aspergillus fumigatus genes that are  
 XX essential and are potential targets for drug screening. The methods and  
 XX compositions of the present invention are useful for diagnosing and/or  
 XX treating invasive Aspergillus fumigatus infection, including the allergic  
 XX forms of the disease, such as Farmer's lung disease. They can also be  
 XX used in various drug discovery purposes, such as expression of the  
 XX recombinant protein, hybridization assay and construction of nucleic acid  
 XX arrays. The present sequence represents an Aspergillus fumigatus  
 XX essential gene protein sequence, used during diagnosis and drug  
 XX development in the invention. These genes share a high degree of sequence  
 XX conservation with known essential genes of candida albicans. The sequence  
 XX data for this patent is not represented in the printed specification, but  
 XX was obtained in electronic format from WIPO.  
 XX SQ Sequence 845 AA;

Alignment Scores:  
 Pred. No.: 6.25e-54 Length: 845  
 Score: 842.50 Matches: 230  
 Percent Similarity: 42.7% Conservative: 96  
 Best Local Similarity: 30.1% Mismatches: 254  
 Query Match: 23.3% Indels: 183  
 DB: Gaps: 21

US-10-620-914-44 (1-1947) x ADR86056 (1-845)  
 QY 124 AAGAAGGCGC---GATGATCACGCTGCTCGCTGGAGAGCTTCTACGGCGCCCGCCGCT 180  
 Db 69 LysLysGlyLeuAsnGlyGlnGlnAspAlaLeuGluSerPheTyrLysThrGlnAlaGly 88  
 QY 181 GCCTTT----- 186  
 Db 89 ValTyrAspAlaThrArgLysArgLeuLeuCysGlyArgGluAspMetLeuGlyLeuVal 108  
 QY 187 CTGCGCGCGCTG-----GCCGAGCGCTCGAACCTCATCTGG 222  
 Db 109 AlaAlaGlnLeuLysTyrLysValGluAsnLysGluLeuGlnAlaGlyLysAlaIleTrp 128  
 QY 223 GTTGACCTGGGTGGTGGCACTGGGGAGATGTCATGATGATGCTGATTCATCATCGACTG 282  
 Db 129 ValAspIleGlyGlyGlyThrGlyTyrAsnIleGluAlaMetAlaSerPheLeuProVal 148  
 QY 283 GCGAAGTTCAAGTCC----ATCTAGCTGGTGCACCTGTGCCACTCGCTGTGCGAGGTGCC 339



XX Streptomyces viridochromogenes Avi gene cluster polypeptide frame 4.  
 DE Avilamycin; antibacterial; virucide; protozoacide; fungicide; infection;  
 KW medicine; Staphylococcus aureus; biosynthetic gene cluster; enzyme.  
 XX Streptomyces viridochromogenes.  
 OS  
 XX WO200268436-A1.  
 PN  
 XX 06-SEP-2002.  
 PD  
 XX 24-AUG-2001; 2001WO-EP009815.  
 PF  
 XX 25-FEB-2001; 2001DE-01009166.  
 PR  
 XX (COMB-) COMBINATURE BIOPHARM AG.  
 PA  
 XX Weithauer G, Muehlenweg A, Trefzer A, Bechthold A;  
 PI WPI; 2003-018650/01.  
 XX N-PSDB; ABZ37516.  
 DR  
 XX New avilamycin derivatives, useful for treatment of infections, and  
 CC nucleic acid encoding avilamycin synthesis enzymes.  
 CC  
 XX Example 1; Page 68-301; 319pp; German.  
 PS  
 XX The invention relates to avilamycin derivatives (I) with antibacterial,  
 CC virucide, protozoacide and fungicide activity. (I) are useful for  
 CC treatment of infections (bacterial, viral, protozoal or fungal), in human  
 CC or veterinary medicine, particularly where caused by Staphylococcus  
 CC aureus. (I) are more hydrophilic than known avilamycins. The present  
 CC sequence is that of an avilamycin synthesis enzyme from the Streptomyces  
 CC viridochromogenes Avilamycin A biosynthetic gene cluster (ABZ37516)  
 CC ABZ37516)  
 XX  
 SQ Sequence 19938 AA;

Alignment Scores:  
 Pred. No.: 6.77e-25 Length: 19938  
 Score: 462.00 Matches: 254  
 Percent Similarity: 32.1% Conservative: 42  
 Best Local Similarity: 27.6% Mismatches: 257  
 Query Match: 12.8% Indels: 368  
 DB: 6 Gaps: 54

US-10-620-914-44 (1-1947) x ABP76680 (1-19938)

QY 1945 AGTTGTCTCTTGGCGCCCTTCC---GGCGGGCCATGTAGAAGGAGCTGTACATGTTGA 1889  
 DB ThrCysAlaGlyArgArgProGlyArgAla\*\*\*ProAlaArgAlaCys----- 12585

QY 1888 CGCGGTCCATGATGCCCTGAGTGGCGGGCGGATGTCAGCGCAGCTCGAAGCCGCCCTTCT 1829  
 DB ArgGlyGlySerProProArgSerArg---ArgCysArgProArgArgSerCysArgPro 12604

QY 1828 GGA-----TCAGCTCGCGTAGGGCGGGCTGAGGAGGCGGAGCGCCAGATGACGA 1778  
 DB GlyThr\*\*\*ArgSerArgArgArgGlySerHisGlyThrArgAlaArgArgArgPro 12624

QY 1777 TGCGCGCGGGCAACTCTGTTGGCCAGGCACTCGGCGAGCTCGTGGCCACGGGCATAT 1718  
 DB AlaArgAlaGlyArgProAlaGlyCysArg\*\*\*ArgProGlyArgThrProArgArgLeu 12644

QY 1717 CCAGCCAGTCCACGTGTGTCATCAGATCACCT----- 1685  
 DB ProArgLysProGlySerProProArgSerProArgGlnSerArgProProProlePhe 12664

QY 1685 ----- 1685  
 DB ArgProPro\*\*\*TrpAsnProValHisArgArgGluGlnPheTyrTrpProProGlyPro 12684

QY 1684 ---TGGTGTAGTGGCGGCTTTGAGCT-----CCTCCA--- 1655  
 DB ValValTrpProArgTyrAla\*\*\*SerAlaAlaAlaSerSerGlnProArgProProle 12704  
 QY 1654 -----TGA 1652

DB 12705 \*\*\*AlaProSerTyrProThrProCysGlyGlnArgArgGlnileArgSerPheGlnGly 12724  
 QY 1651 AGAAGTTGGTGGAGACGGTCAGGTTGTCACCAACGCCACTCT----- 1610  
 DB AsnSerArgTrpHisArgMetSer---ProProAsnAsnAspArgThrGlyValPhePhe 12743  
 QY 1609 -----TGAGGTGGCGAAGCGG----- 1592.  
 DB ArgPheSerLeuGlyPheArgAlaGlyTrpArgGlyProPheThrProProSerProAla 12763  
 QY 1591 -----CCTCGCGCAGGTAGGTGG 1574  
 DB AspMetSerTyrGlyValSerAlaPheArgThrHisGlyArgProProSerGlyValPro 12783  
 QY 1573 GGCAGTTGTGCGCAGCAACTTGCCTGAGGCGAGTTGTAGTAGAGTAGTTCTCTGCTGC 1514  
 DB \*\*\*GlyValArgSerGlyAspArgArg-----Cys 12793  
 QY 1513 GCAGTGGAGATTCTCCGCCACGCCGTCATGTGTGCGCG-----CGATGTAGTTCT 1463  
 DB Arg\*\*\*TrpTrpAlaAlaSerArgArgCysTrpProAlaArgProGlyArgGlyAlaSer 12813  
 QY 1462 CAATGGGATCGCTCGCCTTGNATCAGCGCGTACTGCTTCCCGCGCAGCGCGCGCGCA 1403  
 DB ProAlaGlyGlyGlyProPro-----ProSerSerProSerArg 12825  
 QY 1402 ACCACAGCAGCGCTTGTGAGAGAGCACCAGGCTCAGAACT-----TGACGAACA 1352  
 DB AlaProAlaProPro-----AlaGlyGlyGluArgAlaGlyProTrpGlyArgVal 12843  
 QY 1351 GCCACACCGCGCTTGGCGCGCTTCTTTCAGAAAGTGGATGAGCATGTTGCTGCCACA 1292  
 DB AlaSerProAlaArgSerAlaArgAlaProArgAla-----Pro 12857

QY 1291 GACGGCGCTCTCTCCATTTGTGGCGCGTGGCGAGCGCTTGACGCTTGTGCCAGTC 1232  
 DB GlyAsnAlaAlaProProAlaThrAspPro-----ArgThrProArg 12872

QY 1231 CCAGCACCGCGCGCAGCACTGCAGCACCGCAGCAGCTTGGCCATCCGCG----- 1181  
 DB ProArgProArg-----ArgArgProAlaArgCysProAlaArgProAlaAla 12889

QY 1180 -----CCTGGTAGTACAGCGCTGTGGAAAGTACCAGAGCGGCT----- 1142.  
 DB ArgProProGlySerAlaGlyArgAlaGlySerProProGlyValProProArgPro\*\*\* 12909  
 QY 1142 ----- 1142

DB 12910 ProGlyArgAlaArgCysArgArgGlyAspArgAlaArgProArgProSerProPro 12929  
 QY 1141 -----TGACCAAGAGTGTGGCTGGTGT----- 1118

DB 12930 ProValAlaCysArgArgArg\*\*\*SerArgArgArgAlaArgProCysGlyAlaAlaGly 12949  
 QY 1117 -----GCGACAGAAAGCGCGCAGCTTCTTCT---CGTACA 1085  
 DB GlyProArgArgArgProAlaArgSerThrSerArgArgProAlaAlaGlyProArgThr 12969

QY 1084 -----GCTCCTCATGCGGGTGCAGCGCTCCGCCAAGAGTGC----- 1043  
 DB ArgArgArgAlaArgAlaProGluArgAlaValHisArgAspArgArgThrValAlaArg 12989

QY 1042 -----ACAGTCTCTCAAACTCCAGCTGCTGTAATGCCACCT 1007  
 DB ArgSerAsnArgAsnArgSerArgValArgProProArgProAlaAlaGlyArgProGly 13009

QY 1006 TCTTCAGCTCCAGAAAGCGCGAGCTGCGCGG---GGTTGCAGTCCACCGACACCTGGC 950

Db 13010 ArgGlyArgProValArgAlaGlyAlaArgProGlyProAlaAlaThrAlaProProAsp 13029  
QY 949 CGG-----CCCCCTGCACAGCAGGTTTCAGGGCAT---TGCAGCCCGCTAGTTCAGGG 899  
Db 13030 ArgArgGlyProAlaArgProSerGlyProArgHisGlyArgAlaArgArgGlySerGly 13049  
QY 898 ----TCACACGGTGTCTTGGGTTGATCTCCATCATCTCCATATCCGGTTCGGGGTCT 842  
Db 13050 ArgArgArgProArgProSerGlyArgProAlaAlaProProArgPro---ArgGlyArg 13068  
QY 841 CCCACGACTCGTGTACAGGAAGTGGCGGCGGACATGGCGGCGCTCCACGCGTCTCT 782  
Db 13069 Pro-----GlySerAlaAlaAspProProSerAlaPro 13079  
QY 781 CGTCAGGCGGTGCCAACGCTGGCGCAGCGCCCAATCCACAGCTAGTAGGGCGCGCA 722  
Db 13080 -----GlyGlySerSerSerCysProGlyGlyArgGly 13090  
QY 721 GCCACGGCAGTGGGATCGAACCTGGGTGTCTCTCCACAGCGCTCCAGCTTCT 662  
Db 13091 AlaLeuArgProArgArgProProGlyThr\*\*\*ProProAlaArgArgProGlyArg 13110  
QY 661 GCTCCAGGTAGCGCGG-----GCTCGGGCGGATCTCAA-----TGT 623  
Db 13111 SerProArgArg\*\*GlyArgGlyCysProGlyThrGlyGlyCysArgArgProAlaCys 13130  
QY 622 TGT-----CGATGT-----CGAAGATCGATC----- 602  
Db 13131 CysAlaThrAlaAlaGlyArgCysProArgArgArgAlaCysProAlaAlaCysGly 13150  
QY 601 GCCAAGAAACGCGCGCACAG-----GCATCT 572  
Db 13151 SerArgArgArgGlyArgAlaArgProLeuAlaProArgArgTrpAenAlaThrAlaGly 13170  
QY 571 GCGCAGGGCGAGTCTACTTGC----- 548  
Db 13171 GlyAlaAlaArgGlyArgAlaAlaGlyLeuArgGlySerArgSerAlaGlyAlaProThr 13190  
QY 547 -----CGCTCAGCTAGAACGCGCACAGCGCGCTCTGGCAGCGGTACGAGC 494  
Db 13191 SerProArgSerArgArgGlySerGlyArgArgProArgArgArgGlyArgSerGlyPro 13210  
QY 493 AAGCTGTGTGATGAGCTTGT-----GGAAACGGTGAATCATCTGTCAGCGAGTAGGAGA 440  
Db 13211 ArgAlaGly-----CysAlaValAlaSerValArgArgThrArgProProArgArg 13227  
QY 439 AGTGATGAGCTGCGG-----TGCCTCAGGGGCGGCAATTGCGAAGCT 392  
Db 13228 ValArgProAlaSerArgGlySerSerAlaArgProArgGlyAlaGlyThrAlaArgArg 13247  
QY 391 CGGCCT-----CCACGACTGGACATTTCTTC----- 365  
Db 13248 ThrProGlyGlyArgProArgProProArg\*\*SerArgArgAlaAlaProGlyCys 13267  
QY 364 -----AGCCTTGGCT-----TCG 350  
Db 13268 AlaGlySerProProSerThrArgArgArgSerProTrpProAlaGlyProArgAla 13287  
QY 349 CTTTCT-----TCTTGCCCA-----CTCGCACA 326  
Db 13288 ProSerCysArgAlaAlaCysAlaValArgTrpAspSerAlaArgArgProArgAla 13307  
QY 325 GCG-----AGTGGCACAGCTCGACCGTAGTGAAGTGAAGT 287  
Db 13308 AlaGlyArgAlaGlyGlyProCysAlaGlyCysGlyArgProArgArg----- 13323  
QY 286 TCGCCAGGTGCGTAAATCAGCCATCATATCGACATTTCTCCAG-----TGCCAC 236  
Db 13324 -----ArgArgProAlaValArgProAlaProAlaProArgArgSerSerGlyArg 13338  
QY 235 CACCCAGGTCAACCCAGATGAGT-----TCGAGCGCTCG----- 200

Db 13339 ArgProSerArgProArgArgGlyHisArgArgSerArgTrpThrSerSerArgThrArg 13358  
QY 199 -----CCAGCGGGCAGCAAAAGGCGCGCTGGGGCCCGT 164  
Db 13359 TrpProArgProProArgHisCysProArgAlaAlaArgArgProArgArgAlaAsp 13378  
QY 163 AGAAGCTCTCCAGCGCAGCAGCGTATCATCGCCCTTCTTGTGCGCGAACCACATATGGC 104  
Db 13379 \*\*\*ProSerProSerArgHisAlaPro---ArgProArgAlaAlaAlaAlaThrTyGly 13397  
QY 103 GCAGAAGCGTCCAGTCTCATCTTCTGAGCTTGTGAGCTTCTCCAGGAGAGTTCT 44  
Db 13398 -----ArgSer-----AlaGlyProArgArgSerArg 13406  
QY 43 TCT 41  
Db 13407 Ser 13407  
RESULT 6  
ABP76679  
ID ABP76679 standard; protein; 19938 AA.  
XX AC ABP76679;  
XX DT 26-FEB-2003 (first entry)  
XX DE Streptomyces viridochromogenes Avi gene cluster polypeptide frame 3.  
XX KW Avilamycin; antibacterial; virucide; protozoicide; fungicide; infection;  
XX KW medicine; Staphylococcus aureus; biosynthetic gene cluster; enzyme.  
XX OS Streptomyces viridochromogenes.  
XX PN WO200268436-A1.  
XX PD 06-SEP-2002.  
XX PF 24-AUG-2001; 2001WO-EP009815.  
XX PR 25-FEB-2001; 2001DB-01009166.  
XX PA (COMB-) COMBINATURE BIOPHARM AG.  
XX PI Weitnauer G, Muehlenweg A, Trefzer A, Bechthold A;  
XX WPI; 2003-018650/01.  
XX DR N-PSDB; AB237515.  
XX PT New avilamycin derivatives, useful for treatment of infections, and  
XX PT nucleic acid encoding avilamycin synthesis enzymes.  
XX PS Example 1; Page 68-301; 319pp; German.  
XX CC The invention relates to avilamycin derivatives (I) with antibacterial,  
XX CC virucide, protozoicide and fungicide activity. (I) are useful for  
XX CC treatment of infections (bacterial, viral, protozoal or fungal), in human  
XX CC or veterinary medicine, particularly where caused by Staphylococcus  
XX CC aureus. (I) are more hydrophilic than known avilamycins. The present  
XX CC sequence is that of an avilamycin synthesis enzyme from the Streptomyces  
XX CC viridochromogenes Avilamycin A biosynthetic gene cluster (AB237515-  
XX CC AB237516)  
SQ Sequence 19938 AA;  
Alignment Scores:  
Pred. No.: 3,198-24 Length: 19938  
Score: 453.00 Matches: 233  
Percent Similarity: 30.7% Conservative: 54  
Best Local Similarity: 24.9% Mismatches: 289  
Query Match: 12.5% Indels: 360  
DB: 6 Gaps: 47  
US-10-620-914-44 (1-1947) x ABP76679 (1-19938)

QY 32 GCTACACCAAGAAAGACTTCTCCCTGGAGAGCTCAAGCTCAGCAGCATGAAGATGACC 91  
Db : : : : :  
6549 AlaArgProArgGlyValAlaAaArgTrpArgTrpSer : : : : : 6560  
QY 92 TGACCGTTCTGCGCATATCTGTTTCG- : : : : :  
Db : : : : :  
6561 : : : : :  
QY 140 ACCTGCTGCTCCCTGGAGAGCTTTCAGCGGCCAGG : : : : :  
Db : : : : :  
6575 : : : : :  
QY 188 CTGCCCGCTGGCGAGCGCTCGAACCTCATCTGGTGGACCTGGTGGTGGCACTGGGG 247  
Db : : : : :  
6593 ThrProAlaArgSerArgArgSerSerAlaAaGlyProThrArgGlyTrpSerTrpGly 6612  
QY 248 AGAATGCGATATGATGGCTGATTACATCGACCTGGCGAAGTTCAAGTCCATCTAGCTGG 307  
Db : : : : :  
6613 GlySer : : : : :  
QY 308 TCGACCTGTCCTGCTGCTGCGAGGTGGCCA : : : : :  
Db : : : : :  
6626 ThrProCysSerThrGlyCysAlaAaArgProProThrSer : : : : :  
QY 341 : : : : :  
Db : : : : :  
6646 ArgArgSerThrArgArgArgSerArgThrArgGlyProArg : : : : :  
QY 380 TCGTGGAGCGCGAGCTTGGCAATTG- : : : : :  
Db : : : : :  
6666 ThrGlyArgArgThrCysAlaAaThrArgCysGlySerProProArgSerProAaArgPro 6685  
QY 434 TCACCTTCTCTCTGCTCAGCATGATTCACCGTTCCACACGTCATCGACAGGCTT 493  
Db : : : : :  
6686 ArgThrAlaAaThrGlyArgSerSerCysProAaArgThrProTrpSerArgThrArgSer 6705  
QY 494 GCTGTGTAACCTGCTCCACAGCGGCTGG- : : : : :  
Db : : : : :  
6706 ThrArgArgTrpProAaAlaAaAlaTrpArgThrCysSerSerAspArgArgTrpGlyAla 6725  
QY 521 : : : : :  
Db : : : : :  
6726 ThrSerProArgProGlyProSerAlaAaProProTrpAlaAaProProThrAlaAaAla 547  
QY 548 GCA : : : : :  
Db : : : : :  
6746 AlaTrpThrGlyArgArgSerSerArgThrGlyThrSerLysSerCysArgProPro 6765  
QY 581 GGTGCGCGCTTCTCTGCG- : : : : :  
Db : : : : :  
6766 GlySerThrAlaAaArgCysGlyValProSerProGlyArgGluArgProSerAlaAaAla 6785  
QY 623 ACATTGACATCGGCGCGCGCGCTGCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 682  
Db : : : : :  
6786 ThrSerThr : : : : :  
QY 683 AGCAGAACCCAGGGTTCCGATCCCTACGTGCGGTGCTGCGCGCCCTCTACTACGTGT 742  
Db : : : : :  
6794 AlaArgProAlaAaSerArgAlaAaThrCys- : : : : :  
QY 743 GGATTGGCGCTGCCCGCTGGCCAGCGCTGGCCAGCGCTGGCAGCGCGCGCGCGCGCGCG 802  
Db : : : : :  
6813 ThrThrAlaAaAaArgThrArgAlaAaAlaThrProSerThr : : : : :  
QY 803 CCATGTTCCCGCCCACTTCTCTGATACGACGAGTGGGAGGACCCCGAGCGCGATATGG 862  
Db : : : : :  
6830 : : : : :  
QY 863 AGGTGATGGATCAACCCCAAGACACAGCGGTGTGACCCCTGACTAGCGCGCGGTGCAATG 922  
Db : : : : :  
6840 : : : : :  
QY : : : : :  
6857

QY 923 CCTGAACCTGCTGGTGCAGG- : : : : :  
Db : : : : :  
6858 ProThrArgCys : : : : :  
QY 959 TGTGCGTGG- : : : : :  
Db : : : : :  
6878 CysGlyTrpProProGlyArgSerArgTrpProProThrArgThrProProGlySerSer 6897  
QY 980 : : : : :  
Db : : : : :  
6898 MetProArgProGlyArgProArgArgThrSerProValTrpArgArgGlyArgCysTrp 6917  
QY 992 : : : : :  
Db : : : : :  
6918 ThrArgAlaSerThrGlySerAsnArgProThrProAlaAaTrpSerProGlyGlyTrpPro 6937  
QY 1016 TTCAGCAGCTGG- : : : : :  
Db : : : : :  
6938 : : : : :  
QY 1058 AGGCGCTGCACCCGCGCATTTGAGGAGCTGTACGAGA- : : : : :  
Db : : : : :  
6956 ValSerValTrpCysThrLeuArg- : : : : :  
QY 1097 AGCTGCGCGCTTCTCTGCT- : : : : :  
Db : : : : :  
6975 CysTrpThrPro : : : : :  
QY 1124 : : : : :  
Db : : : : :  
6995 TrpSerCysThrSerThrArgSerProSerThrAlaAaArgHisArgArgArgArgSer 7014  
QY 1151 : : : : :  
Db : : : : :  
7015 ArgSerArgSerIleProAlaSerProThrProCysThrProTrpSerArgThrGlyArg 7034  
QY 1178 AGGCGCGCATGGCAAGCTGTGCTGGTGTGTCAGTGCCTGGCGCTGGTGGTGGTGGTGG 1237  
Db : : : : :  
7035 AspAlaAaArgTrpArgAlaCysAlaAaCysAlaAaThrArgTrpSerSerArgProArg 7054  
QY 1238 GCAAGACCGTCAAGCGCTCGCAACGCGCCCAAA- : : : : :  
Db : : : : :  
7055 ArgThrProThrSerArgAlaAaAlaThrArgThrArgThrTrpCysArgSerArgValCys 7074  
QY 1295 : : : : :  
Db : : : : :  
7075 ArgArgArgSerCysValProGlyCysSerThrArgCysAlaSerArgSerArgArgArg 7094  
QY 1328 : : : : :  
Db : : : : :  
7095 CysAspSerProProArgProCysThrPheAlaAaArgTrpTrpSerArgGlySerThr 7114  
QY 1361 AGTTGCTGACCTGCTGCTCTTCAAGCGCGTGTGCTGGTGGTGGTGGTGGTGGTGGTGG 1420  
Db : : : : :  
7115 Arg : : : : :  
QY 1421 GCAGCAGTACGCGCTGATCAAGCGCGAGCGCTCCCATTTGAGAACTACATCGCGCGCA 1480  
Db : : : : :  
7135 ProThrSerSerGlySerCysProProSerAlaAaThrSerSerIleThr : : : : :  
QY 1481 CCATGCGCGCGTGGCGGAGAACTCGCACGCTGGCGCAAGCAGAACTACTTCTACTACA- 1537  
Db : : : : :  
7155 SerArgAsnThrArgArgArgThrAlaMetArgProGluArgThrAspSerValThrHis 7174  
QY 1538 : : : : :  
Db : : : : :  
7175 ValSerArgGlyArgArgGlyGluArgAlaAaProAlaGlyProGluThr : : : : :  
QY 1574 CCACCTACTCTGCGG- : : : : :  
Db : : : : :  
7195 ProGluLysHisAlaSerThrIleValValGlyArgArgHisSerMetMetProGlyVal 7214  
QY 1610 : : : : :  
AGAGTG 1615



Db 7215 ProLeuLysArgProAspLeuThrAlaLeuThrAlaArgCysGlyIle\*\*\*ArgArgSer 7234  
 QY 1616 GCGTGGTGGACAACTGACCGTCTCCACCAACTCTTCATGAGAGCTCAAGCGCGCA 1675  
 Db 7235 TyrTrpTrp-----AlaGlyLeuAlaArgGlyCysGlyArgLeu 7247  
 QY 1676 CTTACACAAAGTGAATCTGATGACACACGTGG-----ACTGGC 1714  
 Db 7248 ArgIleProArgProAsnAsnTrpSerArgTrpProIleGluLeuPheSerProMetTyr 7267  
 QY 1715 TGGATA----- 1720  
 Db 7268 TrpIleProSerArgArgProGluTyrArgArgSerGlyLeuSerSerGlyArgProArg 7287  
 QY 1721 -----TGCCCGTGGCCAAACGAGCTGGCGGAGTCCCTGGCCCAAGCAGGTTGCGC 1768  
 Db 7288 ArgArgSerArgLeuProTrpGluThrAlaArgSerSerAlaTrpThrSerProThrSer 7307  
 QY 1769 CGGGCGGCATGTCATCTGGGGTCCGCTCCCTCAGCCCGCCCTAGCCGAGCTATCC 1828  
 Db 7308 SerArgThrProArgAlaGlyTyrProThrProCysSerCysProValThrSer---Ser 7326  
 QY 1829 AGAAGCGCGCTTCGAGTGC----- 1849  
 Db 7327 AlaThrThrThrSerSerCysThrTrpAlaThrThrThrSerTrpThrAlaSerSerThr 7346  
 QY 1850 -----GTCATCGCGCGCGCCACTCAG----- 1873  
 Db 7347 SerTrpArgThrSerAlaAlaThrSerProProArgArgSerCysSerProGlySerPro 7366  
 QY 1873 ----- 1873  
 Db 7367 ThrArgAlaGlySerAlaTrpProAsnLeuThrArgThrAlaGly\*\*\*ProArgSerTrp 7386  
 QY 1874 -----GCTACATGACCGCGCTCAACATGTACA-----GCT 1903  
 Db 7387 ArgSerProArgThrProGlyAlaThrTrpArgSerSerAlaCysThrProSerAlaArg 7406  
 QY 1904 CTTCTTACA-----TGCCCGCGCGAAGCGCGCCCAAGAGGCAACT 1945  
 Db 7407 ProCysThrArgProTrpProAlaSerGlyArgProGlyGlyThrSer 7422  
 RESULT 7  
 ID ABP76682  
 DE ABP76682 standard; protein; 19938 AA.  
 XX AC ABP76682;  
 XX DT 26-FEB-2003 (first entry)  
 XX ST Streptomyces viridochromogenes Avi gene cluster polypeptide frame 6.  
 DE DE  
 DE DE  
 KW Avilamycin; antibacterial; virucide; protozoicide; fungicide; infection;  
 KW medicine; Staphylococcus aureus; biosynthetic gene cluster; enzyme.  
 XX OS Streptomyces viridochromogenes.  
 XX PN WO200268436-A1.  
 XX PD 06-SEP-2002.  
 XX PF 24-AUG-2001; 2001WO-EP009815.  
 XX PR 25-FEB-2001; 2001DE-01009166.  
 XX PA (COMB-) COMBINATURE BIOPHARM AG.  
 XX PI Weithauer G, Muehlenweg A, Trefzer A, Bechthold A;  
 XX WPI; 2003-018650/01.  
 DR N-PSDB; ABZ37516.  
 XX

PT New avilamycin derivatives, useful for treatment of infections, and  
 XX nucleic acid encoding avilamycin synthesis enzymes.  
 PS Example 1; Page 68-301; 319pp; German.  
 XX The invention relates to avilamycin derivatives (I) with antibacterial,  
 CC virucide, protozoicide and fungicide activity. (I) are useful for  
 CC treatment of infections (bacterial, viral, protozoal or fungal), in human  
 CC or veterinary medicine, particularly where caused by Staphylococcus  
 CC aureus. (I) are more hydrophilic than known avilamycins. The present  
 CC sequence is that of an avilamycin synthesis enzyme from the Streptomyces  
 CC viridochromogenes Avilamycin A biosynthetic gene cluster (ABZ37515-  
 CC ABZ37516)  
 XX  
 SQ Sequence 19938 AA;  
 Alignment Scores:  
 Pred. No.: 2,75e-23 Length: 19938  
 Score: 440.50 Matches: 182  
 Percent Similarity: 38.9% Conservative: 60  
 Best Local Similarity: 29.3% Mismatches: 235  
 Query Match: 12.2% Indels: 145  
 DB: 6 Gaps: 29  
 US-10-620-914-44 (1-1947) x ABP76682 (1-19938)  
 QY 155 AGAGCTTCTACGGGCGCCAGCGCGCTTCTGCTGCGCGCTGCGGAGCGCTCGAAC 214  
 Db 5169 ArgArgSerIleSerGlnLeuProValProAspValHisPheTyrProGlnCysAlaAla 5188  
 QY 215 TCATCTGGGTGACCTGGGTGGTGGCACCTGGGGAGAATGCGATATGATGGCTGATTACA 274  
 Db 5189 Thr\*\*\*AsnLeuAsnTyr----- 5194  
 QY 275 TCGACCTGCGAGTCAAGTCCATCTACGTGTCGACCTGTGCCACT-----CGCTGT 328  
 Db 5195 ---GlnTrpGlyAlaProCysProLysLysCysSer\*\*\*SerAlaValGlyArgArgAsp 5213  
 QY 329 GCGAGTGGCCAAAGAAAGAGCGGCAAGGCGCTGGAAGAAATGTCCAGGTGCTGGAGG 388  
 Db 5214 ProArgArgProCysSerAlaArgProAlaCysPro\*\*\*SerTyrTrpAsnGlyArg 5233  
 QY 389 CGACCGCTTGCMAATTGGCGCCCTGAGGCGACCGGAGCGCTCATCACCCTTCTCTACT 448  
 Db 5234 ProSerArgAlaThrThrSerAlaSerArgSerProProArgAlaAlaProSerSerThr 5253  
 QY 449 CGCTCAGCATGATCCACCGTCCCAACGTCACTGACCGAGGCTGTGCTGCTACTGTCCC 508  
 Db 5254 SerSerAlaArgTrpArgSerThrProAlaValThrArgSerArgThrAlaCys--- 5272  
 QY 509 AGACGCGCTGTTGGCGCTTTCGCGACTTCTACGTGAGCGCAAGTACGACCTGCCCTGC 568  
 Db 5273 -----CysCysAlaGly 5276  
 QY 569 GCGAGATGCGTGTGCGCGCTTCTTCTGCGGATCGATCTTCGACATCGACACATTG 628  
 Db 5277 AlaArgArgThrGlyArg-----SerThrGlyArgArgSerSerValPro----- 5291  
 QY 629 ACATCGCGCGCGCGCGCTTCTACGTGAGCAGAGCTGGAGCGGTGTGGAGCAGA 688  
 Db 5292 ThrSerAlaProGlyArgSerThr-----AlaThrThrSerThrCysSer\*\*\*Arg 5309  
 QY 689 ACACCGCGGTTTCGATCCCTACGTGCGCGCTGCGCGCTTCTACTACTAGTGTGGATTG 748  
 Db 5310 ThrArgSerSerArgAlaProThr-----SerProArgAlaSerArgSerSer 5325  
 QY 749 GCGCGCTGCCAGCGTGTGGCGCGCTGACGAGGAGCGGTGGAGCGCGCCCGCTATGT 808  
 Db 5326 AlaCysCysSerThrAlaThrAlaArgLeuGlyArgSerGlyTrpThrProThrProVal 5345  
 QY 809 TCCCGCGCCACTTCTCTGACGCGAGTGTGGAGGAGCGCGCGCGGATATGGAGTGA 868  
 Db 5346 Ser-----TrpSerAlaArgSerSerThrThrSer----- 5355





Db 1970 LeuProProLeuAlaArgProArgPro\*\*\*AlaProArgValProLeuProValArgPro 1989  
|||  
Qy 1858 GGATGC-----AGCGCAGT---CGAAGCCGCT----- 1832  
|||  
Db 1990 GlyThrArgThrProArgProSerAlaArgSerArgProProGlnArgSerArg 2009  
|||  
Qy 1831 ---TCTGATCAGCTCGCGTAGGCGGCTGAGGAGG----- 1796  
|||  
Db 2010 AlaAlaGlyAlaProSerArgThrAlaProAspAlaArgThrProProCysAlaSerPro 2029  
|||  
Qy 1795 -----CGAGCGCCAGATGACGATGCGCGCGCGCAA---CCTGCTTG---CCA 1751  
|||  
Db 2030 \*\*\*SerCysArgAlaArgProArgGlyArgProProArgProProCysThrAlaPro 2049  
|||  
Qy 1750 GGCACCTCGGCA---GCTCGTTGGCCACGCGGCATATCCAGCCAGTCCAGCTGT--- 1700  
|||  
Db 2050 ProThrArgProCysAlaGlyAlaProGlyValArg\*\*GlyArgProAlaGlyCysGly 2069  
|||  
Qy 1699 -----CCATCAGATCACCCTGTGTAGTGGCGCTTTGAGCT 1661  
|||  
Db 2070 AlaAspArgProValArgProProCysGlyThrProTrp-----Ser 2083  
|||  
Qy 1660 CTTCCATGAAGAAGTGGTGGAGCGGTCAAGTGTGTCCACCAGCCCACTCTTGAGGGTGG 1601  
|||  
Db 2084 ProProArgArgAlaArg-g-----TTP 2092  
|||  
Qy 1600 CGAAGCGCGCTCGCGCAGTGTGGGCGCAGTGT---CGCGCAGGAATCTGCGGTGA 1544  
|||  
Db 2093 ProProProAlaArgAlaGly-----GlySerAlaProGlyCysGlyThrProArgArg 2110  
|||  
Qy 1543 GGCAGT-----TGCTAGTAAGTAGTCTGTGTCGCACGTCGAGTCTCCGCCA--- 1493  
|||  
Db 2111 GlySerProTrpAlaTrpArgSerGlyThrAlaArgGlyArg---SerAlaProProThr 2129  
|||  
Qy 1492 ---CGCGTCCAGTGGTGGCGCGAGTGT-----AGTTCTCAATGGGATCCGCTCCGCT 1442  
|||  
Db 2130 ArgArgArgProTrpSerGlyGlySerGlyAspSerSerArgSerGlyProArgPro--- 2148  
|||  
Qy 1441 TGATCAGCGGTACTGCTTCCCGGCACGCGCGCGCCACACAGCAGCGCGCTTGTGA 1382  
|||  
Db 2149 -----ArgAlaCysProProArgArgProArgAlaProGlySerProAlaAla 2165  
|||  
Qy 1381 AGA-----GCACAGGCTCAGCAACTTGCAGCAAGCC 1349  
|||  
Db 2166 ArgArgArgTrp\*\*\*ProArgArgProAlaProSerSerArgTrpArg--- 2183  
|||  
Qy 1348 ACACGAGCGCTGGGCGCTTCTTACGAAGTGGATGACATGTTGCTGT----- 1298  
|||  
Db 2184 ---ProAlaGlyArgGlyArgAlaAlaValSerArgAlaAlaHisCysAlaSerArgSer 2202  
|||  
Qy 1297 -----CCCACAGACGCGCTGCTCCT 1277  
|||  
Db 2203 ProGlySerSer\*\*\*SerAlaArgTrpProAspArgProPro\*\*\*SerGlySerAlaPro 2222  
|||  
Qy 1276 ---CCATTGGGCG-----CGTTGGCAGGCGCTTGAAGCTTGTCCCGAGTCCCA 1229  
|||  
Db 2223 ArgArgLeuGlyAlaArgValArgTrpProGlySer-----SerCysArgAlaPro 2240  
|||  
Qy 1228 GCA-----CCACGCGCAGGC 1214  
|||  
Db 2241 SerArgArgArgGlySerSerGlyAlaGlyArgSerGlyCysArgArgProGly 2260  
|||  
Qy 1213 ACTGCAGCA-----CCACGACAGCTTGC----- 1190  
|||  
Db 2261 SerArgAlaValArgProGlyGlyAlaCysArgArgAlaProGlySer\*\*\*ProAlaArg 2280  
|||  
Qy 1189 CCATGCCGCGCTGTAGTACAGCCGT-----GCTGGAGTACAGAGCGCTTGGACC 1136  
|||  
Db 2281 ProProArgProAlaGlyThrProArgProProArgGlyArgGlyArgGlyProArgAla 2300  
|||  
Qy 1135 AGAAGTTGGCTGGTGGACAGAGGCGCCAGCTTCTCTGT----- 1088  
|||

Db 2301 ArgArgAlaGly---ArgSerThrArgSerAlaAlaArgSerArgProGlyArgPro 2319  
|||  
Qy 1087 ---ACAGCTCTCAATGCGCGGTGCACGCT-----GCCCGACAGCTGCCACA 1040  
|||  
Db 2320 ProArgSerProGlyGlyAlaAlaGlyArgProArgProGlyArgAlaAlaValAlaHis 2339  
|||  
Qy 1039 GTCCTCAAACTCCAGTGTCTGAA-----TGGCCACCTTCTTTCAGCT 998  
|||  
Db 2340 ArgArgGlyArgProAlaArgSerValArgGlySerProProArgProProAlaProVal 2359  
|||  
Qy 997 CCAGAAGCGCGCTGCGCG---GTTGTCAGTCCACCGACACACCTGCGCGCCCT 941  
|||  
Db 2360 ProValArgProAlaProArgThrAlaCysAlaGlyLeuProProAla---ProPro 2378  
|||  
Qy 940 GCACACGAGCTTTCAGGCGCATTCAGCGCGCTAGTCAAGGTCAGCGCTGCTCTTGG 881  
|||  
Db 2379 ArgProAlaAlaGly-----ArgAlaSerSerSerAlaProArgPro--- 2392  
|||  
Qy 880 GGTGTATCTCATCACCTCCATATCCGCTCGGGTCTCCACGACTGGTGTACAGGA 821  
|||  
Db 2393 ---ProArgArgProGlyValProArgGlyProProProProAlaArgAlaAla 2409  
|||  
Qy 820 ---AGTGGCGCGGAACATGGCG---GCC 797  
|||  
Db 2410 ArgArgProValArgTrpSerGlyAlaAlaCysProArgArgCysValaArgCys 2429  
|||  
Qy 796 GCTCCACGCGCTCTCTGTCAGGCGCTGCGCAACGCTGGCGCGCGCCATCCACAGT 737  
|||  
Db 2430 CysProValGlnProArgArgSerAlaGlyArgArgProProSerGlyArgArgSerAla 2449  
|||  
Qy 736 AGTAGG-----GGCGCGCAGCCACCGCATGAGGAGTGC--- 701  
|||  
Db 2450 AlaArgProArgGlyAlaGlyAlaAlaGlyThrSerArgArgAlaProProGlyArgPro 2469  
|||  
Qy 700 ---AACCTGGGTCTCTCTCC--- 680  
|||  
Db 2470 SerGlyThrArgProSerProProProGlyAlaAlaCysProArg\*\*\*GlyProThr 2489  
|||  
Qy 680 ----- 680  
|||  
Db 2490 AlaGlyProProGly\*\*\*ProProAlaArgAlaGlyProValAlaHisGlySerValPro 2509  
|||  
Qy 679 -----ACACGCGCTCCAGTCTTCTGCTCCAGGT 653  
|||  
Db 2510 GlyProProValArgProArgArgValArgAspArgArgProProAlaProPro 2529  
|||  
Qy 652 AGCGCGCGCTCGG--- 638  
|||  
Db 2530 ArgGluAlaAlaArgProGlyProProProProAlaHisSerAlaProSerArgProArg 2549  
|||  
Qy 637 ---GCCCGATGT 629  
|||  
Db 2550 ProSerHisArgValProAlaSerCysHisProGlyGlyArgGlyLeuAspGlyTrpCys 2569  
|||  
Qy 628 CAATGTTGT--- 620  
|||  
Db 2570 AlaGlyCysProAlaThrLeuThrIleGlyArgArgAlaAsnThrCysProLeuTrpVal 2589  
|||  
Qy 619 -----CGATGTCGAAGATCGATCGCCAGAGAAC 590  
|||  
Db 2590 GlySerArgGlyAlaLeuGlyGlnAlaAspArgLeuLysAlaGluIleGlyArgLeuAsn 2609  
|||  
Qy 589 GGC----- 587  
|||  
Db 2610 GlyProValArgSerTrpProAlaProAlaValProAspThrArgSerArgAlaAla 2629  
|||  
Qy 586 ---GGACCGAGGCGATCTCGCGCAGCGGCGAGT--- 557  
|||  
Db 2630 Ser\*\*\*ProThrThrGlyProGlyArgAlaGlyArgAspArgGlyProGlyArgGlyAla 2649  
|||  
Qy 556 ---CGTACTTGGCGCTCA--- 542  
|||  
Db 2650 \*\*\*ArgSerValPro\*\*\*SerArgProProValProAlaTrpArgArgCysArgAlaLeu 2669  
|||

QY 541 -----CGTAGAGTCGGCAACCCACACGAGCGCTCTTGGGACAGGTACG-----ACCAAG 491  
 Db 2670 LeuProArgArgSer-----AlaLeuProProTrrLysArgGly\*\*GlyAlaAspSerAla 2688  
 QY 490 CCGTGGTCGATGACGTTGTGAACCGGTGAATCATCGTGAGCGAGT-----446  
 Db 2689 ProProAsnLeuPheCysAlaAlaGlyGlyLeuSerArgArgSerProSerProTrrPsr 2708  
 QY 445 -----AGGAGAAGGTGATGA 431  
 Db 2709 AlaSerProThrSer\*\*CysArgCysAlaProAspArgGlyAspArgAlaArgGln\*\* 2728  
 QY 430 GCCTCGCGGTCCCTCAGGGGGGCGCAATTTGGACGGTCGGCTTCA-----CGACCTGGA 374  
 Db 2729 ProArgArgArgProThrAlaAlaAlaAlaGlySerArgProProAlaArgProGly 2748  
 QY 373 CATTCCTCCAGCCCTTGGCCTTCGCCCTTCT-----344  
 Db 2749 -----SerAlaProProProSerProGlyArgArgAlaGlySerArgArgArgArg 2766  
 QY 343 ---TCTTGGCCACTCGCACAGCGAGTGGCACAGTGCACACGATAGACTGAACT 287  
 Db 2767 ArgSerArgGlySerGlySerProSerGlyThrGlyArgCysProCysTrrProArgPro 2786  
 QY 286 TCGCCA-----GGTCATGTAATCAG---CCATCATATCGACATTCCTCCCGATGC 239  
 Db 2787 ValProSerArgThrAlaArgCysGlyArgAsnProAlaArgSerAlaPro\*\*Cys 2806  
 QY 238 CAC-----CACCCAGGTCAACCCAGATGAGTTTCGAGCGCTCGG-----200  
 Db 2807 ArgSerArgArgProArgProArgAlaProArgThrGlyProProGlyArgThrSerPhe 2826  
 QY 199 -----CCAGCGCGGCAAAAGGCGAGCGCTCGGCGCCGTAGAGC 158  
 Db 2827 ArgCysThrProArgSerProTrrSerGlnGlySerAlaSerAlaSerValArgArgGlu 2846  
 QY 157 TCTCCAGCGAGACGCTGATCGCCCT 128  
 Db 2847 AlaProGlyArgArgAlaArgHisLeuPro 2856  
 RESULT 9  
 ABP76679  
 ID ABP76679 standard; protein; 19938 AA.  
 XX AC ABP76679;  
 XX DT 26-FEB-2003 (first entry)  
 XX DE Streptomyces viridochromogenes Avi gene cluster polypeptide frame 3.  
 XX KW Avilamycin; antibacterial; virucide; protozoacide; fungicide; infection;  
 XX KW medicine; Staphylococcus aureus; biosynthetic gene cluster; enzyme.  
 XX OS Streptomyces viridochromogenes.  
 XX WO200268436-A1.  
 XX 06-SEP-2002.  
 XX 24-AUG-2001; 2001WO-EP009815.  
 XX 25-FEB-2001; 2001DE-01009166.  
 XX (COMB-) COMBINATURE BIOPHARM AG.  
 XX Weitnauer G, Muehlenweg A, Trefzer A, Bechtold A;  
 XX WPI; 2003-018650/01.  
 XX N-PSDB; AB237515.  
 XX PT New avilamycin derivatives, useful for treatment of infections, and  
 nucleic acid encoding avilamycin synthesis enzymes.

XX Example 1; Page 68-301; 319pp; German.  
 XX The invention relates to avilamycin derivatives (I) with antibacterial,  
 CC virucide, protozoacide and fungicide activity. (I) are useful for  
 CC treatment of infections (bacterial, viral, protozoal or fungal), in human  
 CC or veterinary medicine, particularly where caused by Staphylococcus  
 CC aureus. (I) are more hydrophilic than known avilamycins. The present  
 CC sequence is that of an avilamycin synthesis enzyme from the Streptomyces  
 CC viridochromogenes Avilamycin A biosynthetic gene cluster (AB237515-  
 CC AB237516)  
 XX SQ Sequence 19938 AA;  
 Alignment Scores: 5.03e-23 Length: 19938  
 Pred. No.: 437.00 Matches: 237  
 Score: 32.0% Conservative: 43  
 Percent Similarity: 27.1% Mismatches: 271  
 Best Local Similarity: 12.1% Indels: 324  
 Query Match: 6 Gaps: 51  
 DB:  
 US-10-620-914-44 (1-1947) x ABP76679 (1-19938)  
 QY 1945 AGTTGTCTTCTTGGCGCCCTTCGGCGGCCCATGTAGA-----1907  
 Db 6293 SerThrSerGlyTrrArgPro-----GlyProCysArgGlyProSerValArgThrPro 6310  
 QY 1906 -----AGGAGCTGT-----1898  
 Db 6311 AlaGlyArgLeuCysProValSerAlaThrAlaAlaProTrrProMetTrrSerSerAsn 6330  
 QY 1897 -----ACATCTTTCACGCGGT 1893  
 Db 6331 ArgArgProProArgProProAlaProArgArgSerArgProGluProCysSerArgCys 6350  
 QY 1882 CCATGTAGCCCTCAGTGGCGCGCGCATGACGATGCGCGCGCAACCTGCTTGGCCAGGC 1748  
 Db 6351 ProArgProProArg---ArgArgSerValThrAlaArgProArgSerProSerGlySer 6369  
 QY 1822 GCTCGCGGT-----AGGCG 1808  
 Db 6370 ArgArgAlaLeuThrTrrProArgSerGlyThrProTrrProIleGlySerProArgTrr 6389  
 QY 1807 GGCTGAGGAGCGGCGCGCGCATGACGATGCGCGCGCAACCTGCTTGGCCAGGC 1748  
 Db 6390 SerThrGlyArgProSer-----TrrArgProAlaGlyThrAsnTrrPro---6404  
 QY 1747 ACTCGGCCA-----GCTCGTGGCCACGCGCATATCCAGCCAGTCCACGTGGTCCATCA 1694  
 Db 6405 -----ProGlyCysAlaArgTrrProLeuArgAsnProArg-----ProVal 6418  
 QY 1693 GAATACCTTGGTGTAGTGGCGCTTTGAGCT-----1661  
 Db 6419 SerSerProGlyProArgCysArgMetProAlaValProSerGlyCysSerArgValThr 6438  
 QY 1660 ---CCTCATGAAGAATTGG-----TGGAGCGGTGAGTGGCGCGCAGTTCACACACGC 1616  
 Db 6439 GlyProSerGlyProGlyTrrAlaAlaAsnCysTrrLysProSerArgSerProArg 6458  
 QY 1615 CACTCTTGGGTGGCGAAGCGCGCTTCGCGCAGTAGTGGGCGCAGTTCGCGCAGGA 1556  
 Db 6459 \*\*\*SerThrAsnTrrSerArgCysSerArg-----ArgArgSerGlySerArgProGly 6476  
 QY 1555 ACTTGGCGGTGAGCGACTTGTAGTAGAAGTAGTCTCTCTCGCGCAGTGGCGAGTTCCTCG 1496  
 Db 6477 ArgCys-----CystTrrArgAlaThrThrPro 6485  
 QY 1495 -----CCACGCGGTCCATGTTGGCGCGAGTAGTGTCT-----1463  
 Db 6486 ArgSerThrAlaProArgPro\*\*\*SerSerArgCysSerSerGlySerProArgCysGly 6505  
 QY 1462 ---CAATGGGATGCCCTCGCTTGTATCAGCGGTACTGCTTCCCGCGCACGCGCGC 1406



PF	24-AUG-2001; 2001WO-EP009815.	QY	479	TCATCCAGCAGGCTT---GCTCGTACCTGTCCCAAGACGCGCTGTGGCGGTTGCCGACT	535
XX					
PR	25-FEB-2001; 2001DE-01009166.	Db	8715	AlaSerThrSerAlaAlaArgSerAlaIle***AlaSerAlaThrThrProAsp	8734
XX					
PA	(COMB-) COMBINATURE BIOPHARM AG.	QY	536	TTTACGTGAGCGGCAAGTACGACCTGCCCTGCGCCAGATGCCCTGGTCCGCCGCTTCT	595
XX					
PI	Weitnauer G, Muehlenweg A, Trefzer A, Bechthold A;	Db	8735	ArgThrThrAspAlaTrpArg-----	8742
XX					
DR	WPI; 2003-018650/01.	QY	596	TCGTGGCGATCGATCTTCGACATCGACACATTCGACATCGCCCGCGAGCGCGCGCT	652
DR	N-PSDB; ABZ37515.				
XX		Db	8743	PheArgAspArgThrCysSerProThrThrProArgTrpArgArgCysAlaArgProAsp	8762
PT	New avilamycin derivatives, useful for treatment of infections, and				
XX	nucleic acid encoding avilamycin synthesis enzymes.	QY	653	ACCTGAGCAGCAAGCTCGAGCGCG-----TGTGGGAGCAGA	688
XX					
PS	Example 1; Page 68-301; 319pp; German.	Db	8763	ProTTPheArgSerArgSerAlaArgAlaThrTTPCysThrThrSerGlySerSer	8782
XX					
CC	The invention relates to avilamycin derivatives (I) with antibacterial,	QY	689	ACACCCAGGTTGATCCCTACGTGCGTGGC-----	721
CC	virucide, protozoacide and fungicide activity. (I) are useful for				
CC	treatment of infections (bacterial, viral, protozoal or fungal), in human	Db	8783	ThrValArgAlaArgThrProThrAlaValGlyGlyGlyProSerSerSerThrPro	8802
CC	or veterinary medicine, particularly where caused by Staphylococcus				
CC	aureus. (I) are more hydrophilic than known avilamycins. The present	QY	721	-----	721
CC	sequence is that of an avilamycin synthesis enzyme from the Streptomyces	Db	8803	ThrArgProHisArgArgArgProGluArgCysAlaAlaGluArgGlnAlaAspIleGlu	8822
CC	viridochromogenes Avilamycin A biosynthetic gene cluster (ABZ37515-				
CC	ABZ37516)	QY	722	-----TGGCGCGCCCT-----	733
XX					
SQ	Sequence 19938 AA;	Db	8823	ArgGlnValArgGluSerPheGly**GlyLeuCysSerProProArgThrGlyArgAsp	8842
	Alignment Scores:	QY	734	ACTACGTGTGGATTGGCGCGCTCCAGCGCTGGCCACGCC-----TGCACGAGAGCGCG	790
	Pred. No.: 6,52e-23 Length: 19938				
	Score: 435.50 Matches: 230	Db	8843	IleThrCysAlaTrpCysProTTProGlyArgCysAlaProArgAspThrArgSerGly	8862
	Percent Similarity: 33.7% Conservative: 49				
	Best Local Similarity: 27.8% Mismatches: 201	QY	791	TGG-----AGCGCGCGCCCATGTTCCCGCCACCTTCTCTGTACA	829
	Query Match: 12.0% Indels: 347				
	DB: 6 Gaps: 53	Db	8863	TrpArgAlaArgArgArgArgAenAlaGlySerPro-----ArgProAla**Cys	8879
	US-10-620-914-44 (1-1947) x ABB98398 (1-19938)	QY	830	CGCAGTCGTGGAGAGACCCCGAGCGGATATGAGG-----TCATGAGATCAACCCCA	883
		Db	8880	ArgSerArgCysSerThrArgArgThr**TrpLysAlaProGlyTTPArgSer	8897
		QY	884	AGGACAGGTGTGACCTGACTAGCGGGCTGCAATGCCCTGAACCTCTGGTGC	940
		Db	8898	-----MetCysArgPro-----CysThr	8903
		QY	941	-----AGGGCGCGCGCAGGTGG-----TGTCTGTGGACTGCAACCCCGCAGTCGCGCG	991
		Db	8904	HisArgArgSerProValArgGlyArgCysArgCysThrArgSerProGlySerArg	8922
		QY	992	TTCTGGAGCTGAAGAAGGTGGCCATTTCAGCAGCTGGAGTTTGAGGACGTGTGGCAGTGT	1051
		Db	8922	-----	8922
		QY	1052	TCGGCGAGGGCGTGCACCCCGCATTTGAGGAGCTGTACGAGAGAAGACGTGGCGCCTTCC	1111
		Db	8923	-----CysSerArgSerThrThrSer-----	8929
		QY	1112	TGTGCAAAACAGCCACA-----ACTTCTGTCTCAAGCGCTCTGTACTTCCAGCAGC	1165
		Db	8930	-----ThrProAlaThrCysAlaThrSerGlyArgSerArgSerThrArgCysSerAla	8947
		QY	1166	GCCTGTACTACCAAGCGCGCATGGCAAGCTGTCTCTGG-----TGCTGCAGT	1213
		Db	8948	AlaMetThrThrArg***AlaSerAlaThrThrGlyGlyProThrTTPCysThrThr	8967
		QY	1214	GCCTGGCGGTGTCTGGGACTGGCA-----	1240
		Db	8968	SerTrpProSerArgAlaHisTrpSerProArgSerGlyAlaCysArgAlaSerMetCys	8987
		QY	1241	AGACCGTCAAGCGCTCGCCAAACCGCCCAAA-----TGG-----	1276
		Db	8988	ArgProGlySerSerAlaProTrpArgProSerArgGlySerThrTrpSerArgThr	9007
		QY	1277	-----AGGACGACGCGCTGTGTGGGACAGCAACATGCTCATCCACTTCG	1321

PF 24-AUG-2001; 2001WO-EP009815.  
 XX 25-FEB-2001; 2001DE-01009166.  
 XX (COMB-) COMBINATURE BIOPHARM AG.  
 PA Weitnauer G, Muehlenweg A, Trefzer A, Bechthold A;  
 XX WPI; 2003-018650/01.  
 DR N-PSDB; ABZ37515.  
 XX New avilamycin derivatives, useful for treatment of infections, and  
 PT nucleic acid encoding avilamycin synthesis enzymes.  
 XX Example 1; Page 68-301; 319pp; German.  
 PS The invention relates to avilamycin derivatives (I) with antibacterial,  
 CC virucide, protozoacide and fungicide activity. (I) are useful for  
 CC treatment of infections (bacterial, viral, protozoal or fungal), in human  
 CC or veterinary medicine, particularly where caused by Staphylococcus  
 CC aureus. (I) are more hydrophilic than known avilamycins. The present  
 CC sequence is that of an avilamycin synthesis enzyme from the Streptomyces  
 CC viridochromogenes Avilamycin A biosynthetic gene cluster (ABZ37515-  
 CC ABZ37516)

SQ Sequence 19938 AA;

Alignment Scores:  
 Pred. No.: 6,52e-23 Length: 19938  
 Score: 435.50 Matches: 230  
 Percent Similarity: 33.7% Conservative: 49  
 Best Local Similarity: 27.8% Mismatches: 201  
 Query Match: 12.0% Indels: 347  
 DB: 6 Gaps: 53

US-10-620-914-44 (1-1947) x ABB98398 (1-19938)

QY 23 GGCTCGGAGCTACA-----CCAGAAGA-----ACTTCTCCCTGGAGAAGCTCA 67  
 Db 8581 GlyGluArgArgHisPro\*\*\*ProArgHisTrpThrGlyProArgArgSerSer 8600  
 QY 68 AGCTCAGCAGCATGAAGGATGACCTGACCGCTTCTGCGCCCATATGTGTTCGCGCAGCA 127  
 Db 8601 GlySerAsp-----GlyThrAlaSerArg 8608  
 QY 128 AGGCGGATGATCAGCTGCTGCGCTGGAGAGCTTCTACGGGCGCCAGG-----CCG 178  
 Db 8609 Thr-----ArgAlaArgPheSerProArgThrArgSerPro 8620  
 QY 179 CTG-----CCTTTGCTGCGCGCTGCGCGGAGCGCTCGA 211  
 Db 8621 ValSerArgArgAlaProAenGlySerSerProAlaSerProThrArgAlaCysAlaPro 8640  
 QY 212 ACCTCATCTGGGTTGACCTGGGTGGTGGCACTGGGAGAAATGTCGATATGATGCTGATT 271  
 Db 8641 ThrThrThrGly----- 8644  
 QY 272 ACATCGACCTGGCGAAGTTCAAGT---CCATCTACGTGGTGGACCTGTGCCACTCGCTGT 328  
 Db 8645 ThrSerProSerArgAlaMetSerValProSer-----SerThrAlaCys 8659  
 QY 329 GCGAGGTGGCCAGAGAGCGCAAGGCGCAAGGCGCTGGAAGA-----ATGTCAGGTGC 382  
 Db 8660 ThrThrTrp---ArgSerArgThrGlyProArgAenGlyThrCysCysThrAlaArgAen 8678  
 QY 383 TGGAGGCGCAGCTTGCCAAATTTGCGCCCTCGAGGCGCAGCGACGCTCATCACCTTCT 442  
 Db 8679 TrpArgSerSer-----ProProArgSerTrpThrAlaArgSerCysProPro 8694  
 QY 443 CTTACTCGCTCA-----CGATGATTCGACCGTTCACAAAG 478  
 Db 8695 ProThrArgSerSerArgSerArgThrGlyProArgCysHisGlyThrValThr 8714

```

Db 9008 ArgCysArgAlaSerArgSerThrAlaSerSerGlyAlaAlaThrGlySerSerThr--- 9026
QY 1322 TGAAGAACGGGCCA---AGCGCTGG---TGTTGGCTGT 1354
Db 9027 --ArgTrpThrProHisArgThrTrpArgCysArgArgTrpAlaThrArgCysGlyCys 9045
QY 1355 TCCTCAAGTTCTGAGCCTGCTCTTCAACAGGCGC----- 1393
Db 9046 ProCys-----AlaThrCysProThrThrAlaProAlaArgThrArgGlySer 9062
QY 1394 -----TGCTGTGGTTCGGCGCGCGCTGC-----CGGCAAGC 1426
Db 9063 TrpGlyArgSerGlyAlaAspAlaCysAlaSerSerGlyAlaThrArgProArgAlaSer 9082
QY 1427 AGTACGGCTGATCAAGCGGCGGCGCATCCCA---TTGAGAACTACATCGCGCA--- 1480
Db 9083 SerAlaProThrThrArgArgCysGlyThrProSerThrArgArgCysGlyArgAlaArg 9102
QY 1481 -----CCATGGACGGCGTGGCGGAGAACTCGCACGTGC 1513
Db 9103 ArgTrpCysSerArgArgArgAsnArgTrpArgSerTrpGluArgCysArg----- 9120
QY 1514 GCAAGCGAACTACTTACTACACTGCTCAACGGCAAGTTCTCGCGGCACTGCTGC 1573
Db 9121 -----ProAlaSerGlyCysCysAlaThrAla 9129
QY 1574 CCACCTACCTCGCGGCGGCTTGCACACCTCAAGAGTGGC---TGGTGCGACAAC 1630
Db 9130 -----ArgTrpSerSerSerArgThrAlaThrCysTrpCysThr 9143
QY 1631 TGACCGTCTCCACCA-----ACTTCTTCTGAGGAGCTCAAGCGCGCACCTACA 1681
Db 9144 ThrAlaAlaProThrAlaThr---ThrAlaSerTrpArgGlyPheArgSerCysProSer 9163
QY 1682 CCAAGGTGATCTGTAGGACCACTGAGTGGCTGCGTGGATATGCGCGTGGCCACGAGCTGG 1741
Db 9164 Pro-----SerThrThrThr----- 9168
QY 1742 CCGAGTGCCTGGCCAGAGTGGCGCGGCGGCGCATCGTCATCTGGCGCTCCGCT--- 1798
Db 9169 -----Arg---SerAlaAlaAlaValSerIleProProAla 9180
QY 1799 -----CCCTCAGCGCGCTTACCGCGAGCTGATCCAGAGCGCGGCTTCG 1843
Db 9181 ProPro---ProCysProAlaLeuArgProArgProArgSerArgArgArgCysAla 9200
QY 1844 ACCTGC-----GCTGCATCCGCGCGCCACTCAGGCTTACATGACCGCGTCAACA 1894
Db 9201 GlyCysCysSerThrThrAlaThrAlaGlyArgArgAsnCysGlyThrAlaSerSer 9220
QY 1895 -----TGTCAGCTCTTCTACATGTCGCGCGGA 1924
Db 9221 GlyHisLeuArgProProArgSerProGlyCys---ProAlaSerTrpArgThrAla 9239
QY 1925 AGGGCGCAAGAGGCAACT 1945
Db 9240 GlyCysArgArgArgThrSer 9246

RESULT 11
ID ABP76678
ABP76678 standard; protein; 19938 AA.
XX AC ABP76678;
XX DT 26-FEB-2003 (first entry)
XX DE Streptomyces viridochromogenes Avi gene cluster polypeptide frame 2.
XX DX
XX XX Avilamycin; antibacterial; virucide; protozoacide; fungicide; infection;
XX KW medicine; Staphylococcus aureus; biosynthetic gene cluster; enzyme.
XX XX Streptomyces viridochromogenes.

```

```

XX WO200268436-A1.
XX 06-SEP-2002.
XX 24-AUG-2001; 2001WO-EP009815.
XX 25-FEB-2001; 2001DE-01009166.
XX (COMB-) COMBINATURE BIOPHARM AG.
XX Weitnauer G, Muehlenweg A, Trefzer A, Bechthold A;
XX WPI; 2003-018650/01.
XX N-PSDB; AB237515.
XX New avilamycin derivatives, useful for treatment of infections, and
XX nucleic acid encoding avilamycin synthesis enzymes.
XX Example 1; Page 68-301; 319pp; German.
XX The invention relates to avilamycin derivatives (I) with antibacterial,
XX virucide, protozoacide and fungicide activity. (I) are useful for
XX treatment of infections (bacterial, viral, protozoal or fungal), in human
XX or veterinary medicine, particularly where caused by Staphylococcus
XX aureus. (I) are more hydrophilic than known avilamycins. The present
XX sequence is that of an avilamycin synthesis enzyme from the Streptomyces
XX viridochromogenes Avilamycin A biosynthetic gene cluster (AB237515-
XX AB237516)
XX SQ Sequence 19938 AA;
XX
XX Alignment Scores:
XX Pred. No.: 1-72e-21 Length: 19938
XX Score: 416.50 Matches: 228
XX Percent Similarity: 33.5% Conservative: 40
XX Best Local Similarity: 28.5% Mismatches: 248
XX Query Match: 11.5% Indels: 283
XX DB: 6 Gaps: 52
XX
XX US-10-620-914-44 (1-1947) x ABP76678 (1-19938)
QY 1939 CTTTCTTGGCGCTTCCGCGGCGCATGTAGAAGAGCTGTATCATTTGACCGGTCCA 1880
Db 17422 ProSer---ArgProTrpGlyArgProTrpAsnArgSerMetSerAsnTrpProGly--- 17439
QY 1879 TGTAGCCCTGAGTGGCGCGC-----GGATGCGAGCGACGTGCAAGCCCGCCT 1832
Db 17440 -----AlaSerTrpArgSerSerArgTrpProGlyCys-----ArgThrSerProArg 17455
QY 1831 TCTGGATCAGCT-----CGGCGTAGCGGCGCTGAGGAGCGGAGC----- 1790
Db 17456 AlaGlyCysAlaTrpGlyArgTrpArgAlaGlySerGlyArgSerAlaArgCysSer 17475
QY 1789 -----GCCAGATGACGATCCGCGCGCACTGTCTTGGCCAGGCACTCGGCCA--- 1739
Db 17476 ThrProAlaArgThrGlyCys-----AlaSerSerAlaGlySerGlySerArgProPro 17493
QY 1738 -----GCTCGTTGG-----CCAGCGCATAT---CCAGCCAGTCCA--- 1706
Db 17494 ThrArgArgTrpProProSerGlySerProArgLeuTrpGlyThrAlaSerProTrp 17513
QY 1705 -----CGTGTCTCCATCAGAATCACCT---TGG-----TGTAGTGC 1673
Db 17514 ThrSerThrTrpThrGlyProSerProGlyProProHisTrpSerSerArgCysThrCys 17533
QY 1672 GCGCTTTGAGTCTCTCCATGAAGAAAGTTGGTGAGACGGTGTTCACCCAGCCAC 1613
Db 17534 -----TTPTrpArgAla-----ProArgHis 17541
QY 1612 TCTTGGGCGTGGCAAGGCGC-----CCTCGC 1586
Db 17542 AlaGlyProSerArgArgAlaGlyArg---ArgArgSerSerAlaArgCysTrpProArg 17561

```



QY 1585 GCAGGTAGTGGGCGAGTCTCTCGCAGGAACTTGGCGGTGAGGCGAGTTGTAGTACA--- 1529  
 Db |||||  
 QY 17562 ArgGlyArg-----ArgGlnAlaGlyCys-----AlaCysCysSerArgGly 17575  
 Db |||||  
 QY 1529 ----- 1529  
 Db 17576 GlyProHisGlyAspGlyArgTrpArgProTrpArgArgSerAlaHis\*\*\*ProSerArg 17595  
 QY 1528 ---AGTAGTTCTGTCGCGACGTCGAGTTCTCCGCCCGCTCCATGTCGCGCGCA 1472  
 Db |||||  
 QY 17596 ProThrThrValCysHisArgCysSerProArgProArgSerThrCysCysAlaAla 17615  
 QY 1471 TGTAGTTCTCAATGG-----GGATCCCGCTCCGCTTATGATCGCGCGT 1430  
 Db |||||  
 QY 17616 ArgAlaProArgTrpProProTrpSerThrSerSerCysAlaProPro-----SerThr 17633  
 QY 1429 ACTGCTTGGCGGACGCCGCCCGGAAACACAGCAGCGCTTGTGTAAGAGCACCA--- 1373  
 Db |||||  
 QY 17634 ThrAlaCysSerProSerSerArg-----AlaAlaProProThrProArgArgProArg 17651  
 QY 1372 -----GGCTCAGCACTTGACGAACAGCCACA 1346  
 Db |||||  
 QY 17652 GlySerCysArgThrArgTrpProSerGlySerGlyProArgArgSerThrArgThr 17671  
 QY 1345 CCAGCGGCTTGGCGGCTTCTTCCAGCAAGTGGATGAGCATGTTGCTGCCACAGACGGC 1286  
 Db |||||  
 QY 17672 AlaGlySerHisCysAlaAlaThrArgProSer---AlaProAlaSerProSerProSer 17690  
 QY 1285 GCTGCTCTCCATGTGGCGCGCTTGGCAGGCGCTTGACGGTCTTGGCCAGTCCACGCA 1226  
 Db |||||  
 QY 17691 AlaGlyArgSerAlaThrArgTrpGlySerAlaAlaArgSerArgProSerProHis 17710  
 QY 1225 CCAGG-----CCAGGCACTGCAGCACCCAGC-----ACAGCTTGC 1190  
 Db |||||  
 QY 17711 SerArgThr\*\*\*ProProArgGlyArgAlaThrSerTrpSerValArgProAlaCys 17730  
 QY 1189 -----CCATCGCCCTCTGGTAGT-----ACA 1169  
 Db |||||  
 QY 17731 \*\*\*SerThrGlyProGlyArgProArgSerAlaValCysCysAla\*\*\*HisArgCysThr 17750  
 QY 1168 GGCGGTGCTGGAAGTACAGAGCGCTTGGACCAAGAGTTGTGGCTTGGCAGGGA 1109  
 Db |||||  
 QY 17751 AlaProThrGlyArgThrArgSer-----ProThrGly 17761  
 QY 1108 AGGCGCGCAGTTCTCTGTACAGTCTCTCAATGCGCGGTGCACGCCCTCCCGCAACA 1049  
 Db |||||  
 QY 17762 ArgGlyGlyAlaAlaSerArgSerProSerThrArgCysGlyAlaSerArgArgPro 17781  
 QY 1048 GCTGCC-----ACAGTCTCTCAAACT----- 1028  
 Db |||||  
 QY 17782 GlyAlaAlaArgTrpAlaCysSerSerTrpThrArgProThrThr\*\*\*ArgIleProGly 17801  
 QY 1027 CCAGTCTGTAAGTGGCCACCTTCTTTCAGCTCCAGAGCGCGAGTGGCGGGTTCAGT 968  
 Db |||||  
 QY 17802 ProAlaAlaProTrpProSerProSerGlyArgSerThrAlaThrAla---CysCysSer 17820  
 QY 967 CCACCGACACACTGGCGCGGCCCTTGCACAGCAGGTTCAGGGCATTCGACGCCGCGC 908  
 Db |||||  
 QY 17821 \*\*\*ProValArgProTrpArgThrAlaSerArgSerSerAlaAlaTrpCys----- 17837  
 QY 907 TAGTCAGGGTCAGCACCGTGTCTTGGGGTTGATCTCCATCACCTCCATCCGCTCGG 848  
 Db |||||  
 QY 17838 -----GlySerSerProSerTrp-----Arg 17845  
 QY 847 GGTCTCCACGACTGGTGTACAGGAAGTGGCGGGAACATGGCGCGCTCCACGC 788  
 Db |||||  
 QY 17846 ArgProSerThrSer-----ThrThrAlaTrp-----Arg 17855  
 QY 787 GCTCTCTGTGAGGGCGTGGCCAAAGCTGGGCGAGCGGCCCATCCACAGCTAGTAGGGG 728  
 Db |||||  
 QY 17856 AlaProArg-----ArgSerAlaArgSerHis 17865

QY 727 CGC-----GCAGCCACGGCAGCATAGGGGATCGAACCCCTGGGTGTTCTGTCTCCACA 677  
 Db |||||  
 QY 17866 ArgSerIleCysAlaAlaThrSerArgThrSerSerProAsnSerArgArgSerThr 17885  
 Db |||||  
 QY 676 CGCGCTCCAGCTTCTCTCAGGTAGCGCGCGCTCGGGCCGATGTCAATGTTGTGCA 617  
 Db |||||  
 QY 17886 ArgThrSerGlyArgSerProAlaArgArg----- 17895  
 QY 616 TGTGCAAGATCATCTCCAGAAACACGGCGGACCCAGGGCATCTCGCGCAGGGCAGGT 557  
 Db |||||  
 QY 17896 -----ThrArgArgProThrAlaArgProCysAlaProAlaThrSerTrp 17910  
 QY 556 CGTACTTGGCGCTCAGTGAAGTCCGCAACGCCACACCGCGCTCTTGGGACAGGTACG 497  
 Db |||||  
 QY 17911 Arg---CysAlaGlyArgArgThrArgValArgSerGlyProArgArgCysThrGlyCys 17929  
 QY 496 AGCAACGCTGGTGCATGACGTTGTGGAACGGTGAATCATCGTGAGCGAGTAGGAAGG 437  
 Db |||||  
 QY 17930 AlaArg----- 17931  
 QY 436 TGATGAGCGTCCGCGTCCCTCAGGGGCGCAATTGGCAAGCGTCCGCTCCACGACCT 377  
 Db |||||  
 QY 17932 -----SerIleuArgArgProProArgThr 17939  
 QY 376 GGAATCTTCCAGCCCTTGGCCTTCGCTTCTCT-----TGGCCA----- 335  
 Db |||||  
 QY 17940 Gly\*\*\*ArgSer-----TrpCysSerProArgSerGlyThrTrpTrpGlnTrpAla 17957  
 QY 334 -----CCTCGCACGAGTGCACAGTCCGACACAGT---AGATGGACTTGAACCTCG 284  
 Db |||||  
 QY 17958 ArgArgSerArgThrArgProGlyGlnGlyArgGlnAlaArgArgThrSerArgGly 17977  
 QY 283 CCAGGTGCGATGTAATCGACCATCATATCGACATCTCCCGAGTCCACCCAGGTCAA 224  
 Db |||||  
 QY 17978 ProLysArg---GlyGluProGlnArgMetArgGlyAlaArg---ArgTrpLysGlyGln 17995  
 QY 223 CCAGATGAGGTTCCAGCGCTCGG-----CCAGCGCGGAGCAAGAGCAG--- 179  
 Db |||||  
 QY 17996 ProArgArgThrArgArgAlaArgArgSerProProTrpGlyThrHisArgArg 18015  
 QY 178 -----CGGCTGGGCGCGTAGAGCTCTCCAGCGAGCAGCGTGATCAT 134  
 Db |||||  
 QY 18016 ArgSerGlyArgGlyArgAlaGlyCysArgAspArgGlyProGlyValCysSerGlyArg 18035  
 QY 133 CGCCCTTCTTGTGCGCA-----ACCACATATGGCGCAGAACCG 95  
 Db |||||  
 QY 18036 SerArgAlaAlaCysArgProProAlaAspSerSerSerThrThrSerProProArg 18055  
 QY 94 TCAGGTCACTCTTCATGCTGCTGAGCTTTCAGGGAGAGTCTTCTTGTGTGT 35  
 Db |||||  
 QY 18056 ProValThrProTrpCys-----SerArgArgSerArgGlnAlaGluTrpAla 18071  
 QY 34 AGC-----TCGACGCGCGCGCTCAGCCCGGACCCCA 2  
 Db |||||  
 QY 18072 SerThrCysArgProProTrp\*\*\*SerSerAlaSerArgSerSerArgPro 18090  
 Db |||||  
 RESULT 12  
 ID ABP76682 standard; protein; 19938 AA.  
 XX AC ABP76682;  
 XX XX 26-FEB-2003 (first entry)  
 XX DE Streptomyces viridochromogenes Avi gene cluster polypeptide frame 6.  
 XX XX Avilamycin; antibacterial; virucide; protozoacide; fungicide; infection;  
 XX KW medicine; Staphylococcus aureus; biosynthetic gene cluster; enzyme.  
 XX XX Streptomyces viridochromogenes.  
 XX OS  
 XX FN WO200268436-A1.  
 XX XX





QY 517 -----GGCCCTTTGGGACAGGTACGACGAGCTGCTGCATGACGT 476  
 Db 11120 CysLeuAlaLeuGlyGlyAlaAlaLeuArgAlaProAlaValArgProGlyArg----- 11137  
 QY 475 TGTGGAACGGTGAATCATCTGAGCGAGTAGAGAGAGGTGATGAGCGTCCGGTGC--- 419  
 Db 11138 -----ArgThrArg\*\*\*ArgTrpSerAlaAlaAenCysArg 11149  
 QY 418 -----CCTCAGGGGGCGCAATTTGCAAGCTCGGCTCCA---CGACCTGGA 374  
 Db 11150 TrpArgSerGlyProAenArgGlyArgAlaGlyArgAlaProArgProGly 11169  
 QY 373 CATCTCTCCAGCCCTTGGCTTCCGCT-----TCTTCTTGGCCACCTCGC 329  
 Db 11170 -----\*\*\*ProGlyProGluProArgValArgSerArgAlaProSerProArg 11186  
 QY 328 ACAGCGAGTGGCAGGT-----CGACCAGCT 302  
 Db 11187 ArgArgArgGlyAlaGlyAlaIleProGluProProGlyIleArgCysAlaValArg 11206  
 QY 301 AGATGACCTGAACTTCCGAGCTGATGTAAT-----CAGCATCATATCGACATTCT 248  
 Db 11207 Arg-----ArgProGlyArgCysSerAspCysArgProCysAlaGlyArg 11222  
 QY 247 CCCAGTGCACCACCCAGCTCAACCCAGATGAGTTCGAGCGCTCGGCCAGCGGCGAG 188  
 Db 11223 \*\*\*ArgGlyHisGlyAlaMetAlaPro-----ProArgProGlyThrAla 11237  
 QY 187 CAAGGAGCGGCGCTGGGGCCCGTAGAAGCTCTCCAGCGAGCAGCGGTGATCATCGC--- 131  
 Db 11238 Pro\*\*\*GlyArgAlaArgArgTtpAlaArgProGlyArgProArgThrArgArg 11257  
 QY 130 -----CCT 128  
 Db 11258 AlaAlaProValProArgGlyAlaAlaGlyProValProGlyProSerProAlaSerPro 11277  
 QY 127 TCTTGTCTCGGCAACCATATATGGCGCAGAACGGTCACTCTTCATGCTGCTGAGCT 68  
 Db 11278 SerCysAlaArg-----GlyArgGlyArgAspAlaHis----- 11288  
 QY 67 TGAGCTTCTCCAGGAGAGTCTTCTTGTGTGAGTCTCGCAGCGCGC 20  
 Db 11289 -----ArgProArgArgGlySerSerSerArgArgSerAlaGly 11302  
 RESULT 13  
 ABP76678  
 ID ABP76678 standard; protein; 19938 AA.  
 AC ABP76678;  
 CX  
 XX 26-FEB-2003 (first entry)  
 DT Streptomyces viridochromogenes Avi gene cluster polypeptide frame 2.  
 DE Avilamycin; antibacterial; virucide; protozoacide; fungicide; infection;  
 KW medicine; Staphylococcus aureus; biosynthetic gene cluster; enzyme.  
 KW Streptomyces viridochromogenes.  
 XX WO200268436-A1.  
 XX 06-SEP-2002.  
 PD 24-AUG-2001; 2001WO-EP009815.  
 XX 25-FEB-2001; 2001DE-01009166.  
 XX (COMB-) COMBINATURE BIOPHARM AG.  
 PA Weitauner G, Muehlenweg A, Trefzer A, Bechthold A;  
 XX WPI; 2003-018650/01.  
 XX

DR N-PSDB; ABZ37515.  
 XX New avilamycin derivatives, useful for treatment of infections, and  
 PT nucleic acid encoding avilamycin synthesis enzymes.  
 XX Example 1; Page 68-301; 319pp; German.  
 PS The invention relates to avilamycin derivatives (I) with antibacterial,  
 CC virucide, protozoacide and fungicide activity. (I) are useful for  
 CC treatment of infections (bacterial, viral, protozoal or fungal), in human  
 CC or veterinary medicine, particularly where caused by Staphylococcus  
 CC aureus. (I) are more hydrophilic than known avilamycins. The present  
 CC sequence is that of an avilamycin synthesis enzyme from the Streptomyces  
 CC viridochromogenes Avilamycin A biosynthetic gene cluster (ABZ37515-  
 CC ABZ37516)  
 XX SQ Sequence 19938 AA;  
 Alignment Scores:  
 Pred. No.: 3,84e-20 Length: 19938  
 Score: 398.50 Matches: 225  
 Percent Similarity: 33.5% Conservative: 46  
 Best Local Similarity: 27.8% Mismatches: 265  
 Query Match: 11.0% Indels: 273  
 DB: Gaps: 46  
 US-10-620-914-44 (1-1947) x ABP76678 (1-19938)  
 QY 23 GGCTCGGCTACACCAAGAGAACTTCTCCTCGG-----AGAAGCTCAAGC 70  
 Db 17449 GlyCysArgThrSerProArgAlaGlyCysAlaTrpGlyArgTrpArgAlaGlySer 17468  
 QY 71 TCAGCAGCATGAAGGATGACCTGACCGTCTTGGCCATATGTGCTCGGCAAGG 130  
 Db 17469 GlyArgSer-----AlaArgCysSerThrProAlaArg-- 17479  
 QY 131 GCATCATCAGCTGCTGCTGCGAGCTTACGCGGCCCGGAGCGCTGCTGCTGCTG 190  
 Db 17480 -----ThrGlyCysAlaSerSerAlaGlySerGlySerArgPro----- 17492  
 QY 191 CCGGCC-----TGGCGGAGCGCTCGAACCTCATCTGGGTGACCTGGGTGGGCACTG 244  
 Db 17493 ProThrArgArgTrpProProSerGlySerProArgLeu-----Tyr 17507  
 QY 245 GCGAAGATGTCGATATGATGGCTGATTACATCCACCTGCGGAGGTTCAAGTCCATCTACG 304  
 Db 17508 GlyThrAlaSer-----ProTyrThrSerThrTrpThrGlyProSerProGlyPro 17524  
 QY 305 -----TGCTCGACCTGTGCTGCTGCGAGGTGCGCAAGAGCGGCAAGG--- 355  
 Db 17525 ProHisTrpSer-----SerArgCysThrCysTrpTrpArgAlaProArgHis 17541  
 QY 356 -----CCAAGGGCTGAAGAAATGTCAGGTGC----- 382  
 Db 17542 AlaGlyProSerArgAlaGlyArg\*\*\*ArgArgSerSerAlaArgCysTrpProArg 17561  
 QY 382 ----- 382  
 Db 17562 ArgGlyArgArgGlnAlaGlyCysAlaCysCysSerArgGlyGlyProHisGlyAspGly 17581  
 QY 383 ---TGGAGCGCGAGCTTGGCAATTTGCGCCCTGAGGCGACCG----- 424  
 Db 17582 ArgTrpArgProTrpArgArgSerAlaHis\*\*\*ProSerArgProThrThrValCys 17601  
 QY 425 CGAGCTCATCAGCTTCTCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 484  
 Db 17602 HisArgCysSerProArgPro-----ArgSerThrCysCysAla 17614  
 QY 485 ACCAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 544  
 Db 17615 AlaArgAlaProArgTrpProProTrpSerThrSerSerCysAlaProProSerThr 17634  
 QY 545 GCGCAGTAGTACGACCTGC-----CCTGCGCCAGATGC 577

```
Db 17635 AlaCysSerProSerArgAlaAlaProProThrProArgArgProArgGlySerCys 17654
QY 578 CTTGGTGGCGCGCTTTCT---TCTGGCGATCGATCTTCGACATCCAGCAACATTGACATCG 634
Db 17655 ArgThrArgTrpProSerGlySerGlyProArgSerThrThrArgThrAlaGlySer 17674
QY 635 -----GCTCCGAGCGCGCGCTTACC 655
Db 17675 HisCysAlaAlaThrArgProSerAlaProAlaSerProSerProSerAlaGlyArgSer 17694
QY 656 TGGAGCAGAGCTGGAGCGGTGGGAGCAGAACCCAGGCTTCGATCCCTTACGTGC 715
Db 17695 SerAlaAlaThrArgTrpGlySerAlaAlaArgSerArgProHisSerArgThr--- 17713
QY 716 CGTGGCTGCGCGCCCTTACTAGCTGTGGATTGGCGCGCTGCCAGCGTGGCCAGCGCC 775
Db 17714 -----*ProProArg-----GlyArgAlaThrSerTrpTrpSerValArgPro 17728
QY 776 TGCACGAGGAGCGGTGGAGCGCGCCCA-----TGTCCCGCCCACTTCC 823
Db 17729 AlaCys---SerThrGlyProGlyArgProArgSerAlaValCysCysAla---HisArg 17748
QY 824 TGTACAGCAGCTGTGGAGGAGCCCGAGCGGATATGGAGGTGATGG-----AGA 874
Db 17749 CysThrAlaProThrGlyArgThrArgSerProThrGlyArgGlyGlyAlaAlaSerArg 17768
QY 875 TCAACCCCAAGGACACGGTCTGACCTGACTAGCGCGCTGCAATGCCCTCAACCTGC 934
Db 17769 SerProSerThrArgCys-----GlyAlaSerArgPro----- 17781
QY 935 TGTGTCAGGCGCGCGCCAGGTGGTGT-----CGGTGACTGCAACCCCGCGCAGTCGG 988
Db 17782 -----GlyAlaAlaArgTrpAlaCysSerSerTrpThrArgProThrThr***Arg 17798
QY 989 CGCTTCTGGAGCTGAAGAAGTGGCCATTACAGCAGCTGGAGTTGAGGAGCTGTGCAGC 1048
Db 17799 IleProGlyProAlaAlaProTrpProSerProSer-----GlyArg 17812
QY 1049 TGTTCGCGCAGGCGGTGCACCCCGCATGTGAGGAGCTGTACGAGAGAGTGGCGCCCT 1108
Db 17813 SerThrAlaThrAlaCysCys-----Ser***ProValArgProTrpArgThr 17828
QY 1109 TCTGTGCGCAACAGCCCAACTTCTGTGTCAGCGCC-----TCTGTT 1153
Db 17829 AlaSerArgSerSerAlaAlaTrpCysGlySerSerSerProSerTrpArgArgProSer 17848
QY 1154 ACTTCCAGCAGCGCC-----TGTACTACGAGCGCGCATGGCAAGCTGTGCTGGGTGC 1207
Db 17849 ThrSerThrThrAlaTrpArgAlaProArgArgSerAlaArgArgSerHisArgSerIle 17868
QY 1208 TGCAGTGCCTGGCGGTGTGCTGGGACTGGGCAAGACCGTCAAGC-----GCC 1255
Db 17869 CysAlaAla-----ThrSerArgThrSerSerProAlaSerArgArg 17882
QY 1256 TGCACCAACCGCCCAACTGGAGGAGCAGCGCC----- 1288
Db 17883 SerSerThrArgThrSerGlyArgSerProAlaArgArgThrArgProThrAlaArg 17902
QY 1289 GTCTGTGGGACAGCA-----TGCTCATCC-----ACTTCGTGAAGACGGCG 1333
Db 17903 ProCysAlaProAlaThrSerTrpArgCysAlaGlyArgArgThrValArgSerGly 17922
QY 1334 CCAAGCGCTGGTGGCTGTTCGTCAGT-----TCGTGA 1369
Db 17923 ProArgArgCysThrGlyCysAlaAlaArgSerLeuArgArgProProArgThrGly***Arg 17942
QY 1370 GCCTGTGTC-----TCTCAACA 1387
Db 17943 SerTrpCysSerProArgSerGlyThrTyTrpGlnTrpTrpAlaArgArgSerArgThr 17962
QY 1388 AGCCCGTGTGCTGGCGCGCGGTGCGCGCAAGC----- 1426
```

Lu C, Daggett LP;



Db 692 ArgArgThrProThrGlySerGlyProThrTrpArgSerCysSerSerTrpGlyThrGlu 711  
Qy 1157 -----TCCAGCAGCGGCTGTACTACCAAG 1180  
Db 712 ArgHisArgAenTrpArgGlnCysGlySerGlnGlySerAlaArgMetArgArgThrArg 731  
Qy 1181 GCGGCA-----TGGCAAGCTGTCTGGGTCTCAGTGTCTGGCG 1222  
Db 732 AlaAlaSerTrpThrSerThrTrpGluAlaSerSerTrpCysCys-----TrpTrpPro 750  
Qy 1223 TGTGCTGGGAC---TGGCAAGACCGCTCAAGCCG-----1255  
Db 751 TrpGlyTrpProCysTrpSerSerProGlySerThrTrpSerThrGlySerCysAlaThr 770  
Qy 1256 ---TGGCAACGCGCCCAATAGAGGAGGAGCGCCGTC---TGTGGCAGCAGCAATGC 1309  
Db 771 ArgCysProThrHisProSerTrpThrSerCysTrpLeuSerAlaGlyAlaSerThrAla 790  
Qy 1310 TCATCCACTTCGTGAAGACGGGCCA-----1336  
Db 791 AlaSerAlaGlyCysArgAlaSerProAlaHisArgGlyArgProAlaArgThrSerArg 810  
Qy 1337 -----AGCGCTGGGTGGCTGTCTGCTCA-----1360  
Db 811 ProAlaArgProArgProAlaCysSerArgPheCysArgGlnProAlaThrTrpProArg 830  
Qy 1361 ---AGTTCGTAGCCTGTCTCTTCAACAGCGCGTC---TGTGTTCCGGCGCGCG 1414  
Db 831 ArgAlaAlaThrProTrpThrAlaProLeuAlaProSerArgGlyValAlaAlaAla 850  
Qy 1415 ---TGGCGGCAAGCAGTACGGCTGATCAAGCGGACGGCA 1453  
Db 851 ValArgProHisArgProProAlaArgProArgGlyLeuAlaProAlaHisAlaCysPro 870  
Qy 1454 TCCCAATTGAGAATCATCGCGGCACCATGAGCGCGTGGGAGAACT-----1504  
Db 871 ProProThrArgProGlnSerArgAlaProArgAlaGlyAspArgGlnThrGlyValAla 890  
Qy 1505 ---CGCAGTCGCGCAAGC---AGAACTACTTCTACTACAAGTCCCTCACCAGGATCC 1558  
Db 891 ArgArgLeuCysAlaGlyLeuArgSerProArgAlaAlaProArgArgGlyArgPro 910  
Qy 1559 TGGCGGCAACTGCCCACTACTGTC-----1585  
Db 911 CysProThrSer-----ProGluCysArgAlaAlaGlnProGlyArgArgGlyArg 928  
Qy 1586 -----GCGAGCGCGCTTCGCCACCCTCAAGAGTGGCG 1618  
Db 929 CysGlyProGlyThrAlaGlyGlyThrSerArgProProSerGlyProCysArgProArg 948  
Qy 1619 TGTGGCAACACCTGACCG-----TCTCCACCACTTCTTCATGGAGAGCTCAAGAGCGC 1672  
Db 949 AlaValThrThrAlaProPheLeuGluProThrAspProAlaAlaProSerSerArgSer 968  
Qy 1673 GCACCTACACCAAGGTGATCTGATGACACAG-----TGG 1708  
Db 969 SerArgSerProArg-----SerTrpArgThrCysArgCysSerValArgSerSerTrp 986  
Qy 1709 ACTGGC-----TGATA 1720  
Db 987 ProGlyGlyArgProCysThrArgProGlyProGlyAlaArgAlaArgValThrLeuPro 1006  
Qy 1721 TGGCGGTGGCAAGAGCTGGCGAGTGGCTGGCCAGCAGGTGTGGCGCGCGGCGATCG 1780  
Db 1007 CysProAlaProTrpProArgProSerLeuGlyProAlaArgCysProLeuGlyAlaPro 1026  
Qy 1781 TCATCTGGCGCTGGCTCCCTCAGCC---CGCCCTACGCGGAGCTGATCCAGAGCGG 1837  
Db 1027 AlaProProAlaProAlaProThrAlaThrArgProAlaGlyAlaTrpArgArgSer 1046  
Qy 1838 GCTTCGAGCGCTGCTGCTCGCGCGGCGGCGGCTCAGG-----1873  
Db 1047 ArgCysAlaCysArgSerThrGlyArgProAlaArgAlaSerArgGlnGlyProPro 1066

Qy 1874 -----GCTACATGAGCCGCTCAACATGTACAGCTCTCTTCTACA 1912  
Db 1067 ProGlySerThrAspSerThrSerAlaCysThrProThrProThr 1081  
RESULT 15  
AAB26239  
ID AAB26239 standard; protein; 1232 AA.  
XX AAB26239;  
XX 23-FEB-2001 (first entry)  
XX Human N-methyl-D-aspartate receptor subunit NMDAR1A #17.  
XX Human; N-methyl-D-aspartate receptor; NMDA; NMDAR1A; ionotropic;  
KW Glutamate receptor; drug screening; animal model; disease diagnosis;  
KW genetic screening.  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
FH Misc-difference 280 /label= unknown  
FT /note= "encoded by TGA"  
FT  
FT Misc-difference 304 /label= unknown  
FT /note= "encoded by TGA"  
FT  
FT Misc-difference 326 /label= unknown  
FT /note= "encoded by TGA"  
FT  
FT Misc-difference 347 /label= unknown  
FT /note= "encoded by TGA"  
FT  
FT Misc-difference 422 /label= unknown  
FT /note= "encoded by TGA"  
FT  
FT Misc-difference 438 /label= unknown  
FT /note= "encoded by TGA"  
FT  
FT Misc-difference 464 /label= unknown  
FT /note= "encoded by TGA"  
FT  
FT Misc-difference 477 /label= unknown  
FT /note= "encoded by TAG"  
FT  
FT Misc-difference 501.502 /label= unknown  
FT /note= "encoded by TGA"  
FT  
FT Misc-difference 590 /label= unknown  
FT /note= "encoded by TGATGA"  
FT  
FT Misc-difference 712 /label= unknown  
FT /note= "encoded by TGA"  
FT  
FT Misc-difference 744.745 /label= unknown  
FT /note= "encoded by TAG"  
FT  
FT Misc-difference 843 /label= unknown  
FT /note= "encoded by TGATGA"  
FT  
FT Misc-difference 848 /label= unknown  
FT /note= "encoded by TGA"  
FT  
FT Misc-difference 1009 /label= unknown  
FT /note= "encoded by TAA"  
FT  
FT Misc-difference 1152 /label= unknown  
FT /note= "encoded by TGA"  
FT  
FT Misc-difference 1209 /label= unknown  
FT /note= "encoded by TAG"  
FT  
FT Misc-difference /label= unknown  
FT /note= "encoded by TGA"

```

FT Misc-difference 1227
FT /label= UNKNOWN
FT /note= "encoded by TGA; the coding sequence of this
FT protein contains a number of in-frame stop codons which
FT are represented by xaa in the protein sequence"
XX
XX
XX US6111091-A.
XX
XX 29-AUG-2000.
XX
XX 29-SEP-1997; 97US-00940086.
XX
XX 20-APR-1993; 93US-00052449.
XX
XX 20-APR-1994; 94US-00231193.
XX
XX (MERI ) MERCK & CO INC.
XX
XX Daggett LP, Lu C;
XX
XX WPI; 2000-578607/54.
XX
XX N-PSDB; AAA95032.
XX
XX Novel DNA fragment encoding human N-methyl-D-aspartate receptor subunit
XX for identifying mutations and for developing drugs against various
XX disease states.
XX
XX Example 3; Col 255-264; 205pp; English.
XX
XX The present sequence is a subunit (designated NMDARIA) of the human N-
XX methyl-D-aspartate (NMDA) receptor. This is an ionotropic glutamate
XX receptor which contains cation-specific ligand-gated ion channels. The
XX protein and its coding sequence can be used in disease diagnosis and in
XX research to identify other, similar proteins. They can also be used as
XX probes, for example in genetic screening, and in drug screening, as well
XX as enabling the production of animal disease models
XX
XX Sequence 1232 AA;
XX
XX Alignment Scores:
XX Pred. No.: 1.09e-19 Length: 1232
XX Score: 386.00 Matches: 227
XX Percent Similarity: 32.4% Conservative: 76
XX Best Local Similarity: 24.3% Mismatches: 280
XX Query Match: 10.7% Indels: 353
XX DB: 3 Gaps: 53
XX
XX US-10-620-914-44 (1-1947) x AAB26239 (1-1232)
XX
XX 11 GTCGTGACGGCGCGCTCGCGAGCTACACCAAGAAGAACT----- 49
XX 216 ValValThrGluSerTrpArgLeuSerLeuArgGlnLysValArgAspGlyValAlaIle 235
XX
XX 50 -----TTCCTCTGAGAGCTCAAGCTCAGCAGCAGCATGAGATGACC 91
XX 236 LeuAlaLeuGlyAlaHisSerTyrTrpArgGlnHisGlyThrGlnLys----- 251
XX
XX 92 TGACCGTCTCGGCATATGTGGTTCGGCAGACAGAGGGGATGATCAGCTG----- 145
XX 252 -----GlyValCysGlnProArgProGlyThrAlaValPheThrLeuGlyPro 267
XX
XX 146 -----CTCGCTGGAGAGCTTCTACGGC----- 169
XX 268 SerAlaLeuProGlyArgProSerThrGlyThrTyr***MetSerProGlyArgAlaGlu 287
XX
XX 170 CCCAGGCGCTGCTTCTGCTCCCGCCCTGG-----CCGAGC 205
XX 288 ThrSerProSerAlaLeuValGlyThrTrpSerSerProProTrpTrp***SerProSer 307
XX
XX 206 GCTGAACCTCATCTGGTTGACCTGGGTGGGCTGAGGAGAACTCATATGATGG 265
XX 308 ThrGlyThrAlaSerGlyArgTrpTrp-----GlyAlaGlySerMet----- 321
XX
XX 266 CTGATTACATCGACTGCGGAAGTTCAAGTCCA----- 298

```

```
Db 660 GlySerTrpMetProSerSerMetMetLeuLeuSerSerThrThrTrpGlnAlaArgThr 679
Qy 1070 CGCGCTTGGAGGAGCTGTACGAGAAGAGCTGGCGCCCT----- 1108
Db 680 ArgAla-----AlaSerTrpSerProLeuGlyLeuAlaArgSerLeu 693
Qy 1109 -----TCTGTGCGMAACGACCACTTCTGTGTCGA----- 1141
Db 694 LeuProLeuAlaThrAlaSerProCysArgArgThrProThrGlySerGlyPro***Thr 713
Qy 1142 -----AGCGCTCTGGTACT----- 1156
Db 714 TrpArgSerCysSerSerTrpGlyThrGluArgHisArgAsnTrpArgGlnCysGlySer 733
Qy 1157 -----TCCAGCAGCGCTGTACTACACAGCGCGCA 1186
Db 734 GlnGlySerAlaArgMetArgArgThrArg****AlaAlaSerTrpThrSerThr 753
Qy 1187 TGGCAAGCTGTGCTGGGTGCTGCAGTGCCTGGCGGTGGTGGGAC-----TGGGCAAGA 1243
Db 754 TrpGluAlaSerSerThrCysCys---TrpTrpProTrpGlyTrpProCysTrpSerSer 772
Qy 1244 CGGTCAAGCGCC-----TCGCCAAGCGCGCCACATGG 1276
Db 773 ProGlySerThrTrpSerThrGlySerCysAlaThrArgCysProThrHisProSerTrp 792
Qy 1277 AGGAGCAGCGCGCTC-----TGTGGACAGCAACATGCTCATCCACTTCGTGAAGAACGGGC 1333
Db 793 ThrSerCysTrpLeuSerAlaGlyAlaSerThrAlaAlaSerAlaGlyCysArgAlaSer 812
Qy 1334 CCAAGCGCTGTGCTGCTGCTTCAAGTTCGTGGACCTGGTGTCTTCAACAAGGCGC 1393
Db 813 ProAlaHisArgGlyArgProAlaArgThrSerArgPro-----AlaArgProArgPro 830
Qy 1394 TGCTGTGTTGCGCGCGCGTCCGCGCAAGCAGTACGCGCTCATCAAGCGCGACGCA 1453
Db 831 AlaCys-----SerArgPheCysArgGlnProAlaThrTrp***ProArgArgAla*** 848
Qy 1454 TCC----- 1456
Db 849 AlaThrProTrpThrAlaProLeuAlaProSerArgIleGlyValAlaAlaValArg 868
Qy 1457 -----CCA 1459
Db 869 ProHisArgProProAlaArgProArgGlyLeuAlaProAlaHisAlaCysProProPro 888
Qy 1460 TTGGAAGCTACATCGCGCACCACATGACGCGGTGGCGGAGAACT-----CGC 1507
Db 889 ThrArgProGlnSerArgAlaProArgAlaGlyAspArgGlnThrGlyValAlaArgArg 908
Qy 1508 ACGTGGCGCAAGC---AGAACTACTTCTACTACAACTGCCTCAGCGCAAGTTCCTGCGCG 1564
Db 909 LeuCysAlaGlyLeuArgSerProArgAlaAlaProArgArgGlyArgProCysPro 928
Qy 1565 ACAACTGCCCCACTACTGTC----- 1585
Db 929 ThrSer-----ProGluCysArgAlaAlaGlnProGlyArgArgGlyGlyArgCysGly 946
Qy 1586 -----GCGAGCGCGCTTCGCCACCTCAAGAGTGGCGTGGTGG 1624
Db 947 ProGlyThrAlaGlyThrSerArgProProSerGlyProCysArgProArgAlaVal 966
Qy 1625 ACAACTGACCG-----TCTCCACCAACTTCTTCATGGAGGAGCTCAAAAGCGCGACCT 1678
Db 967 ThrAlaProPheLeuGluProThrAspProAlaProSerSerArgSerSerArg 986
Qy 1679 ACACCAAGGTGATCTGATGACACG-----TCGACTGGC 1714
Db 987 SerProArg-----SerTrpArgThrCysArgCysSerValArgSerSerTrpProGly 1004
Qy 1715 -----TCGATATGC 1723
Db 1005 GlyArgProCys***ThrArgProGlyProGlyAlaArgAlaArgValThrLeuProCys 1024
```

```
Qy 1724 CCGTGGCCAAACGAGCTGCGCGAGTGCCTGGCCAAAGCAGGTTGGCCGCGGCGCATCGTCA 1783
Db 1025 ProAlaProTrpProArgProSerLeuGlyProAlaArgCysProLeuGlyAlaProAla 1044
Qy 1784 TCTGGCGCTCGGCTTCCTCAGCC---CGCCCTACGCCGAGCTGATCCAGAGCGGGCT 1840
Db 1045 ProProAlaProAlaProThrAlaThrArgProAlaGlyAlaTrpArgArgSerArg 1064
Qy 1841 TCGACGTGCGCTGCATCCGCGCGCCACTCAGG----- 1873
Db 1065 CysAlaCysArgSerThrGlyArgProAlaArgAlaSerArgGlnGlyProProPro 1084
Qy 1874 ---GCTACATGGACCGCTCAACATGTACAGCTCTTCTTACA 1912
Db 1085 GlySerThrAspSerThrSerAlaCysThrProThrProThr 1098
```

Search completed: March 14, 2006, 01:40:13  
Job time : 460.5 secs



**THIS PAGE BLANK (USPTO)**



```
Qy 943 GGGCCGCCAGGTGGTGTGGTGGACTGCAACCCCGCAGCTGGCGCTTCTGGAGCTG 1002
Db 84 AsnProAlaSerIleAspValValAspLeuAsnProHisIleAlaLeuAsnLysLeu 103
Qy 1003 AAGAAGGTGGCCATTACAGAGCTG---GAGTTTGGAGACGTGTGGAGCTGTTCTGGCGAG 1059
Db 104 LysLeuAlaAlaPheArgHisLeuProAlaHisGlnAspValValArgHisPheGlyArg 123
Qy 1060 GGCGTGACCCGCCATTGAGGACCTGTACAGAAAGAGCTGGCGCCCTCTCTGTCCAA 1119
Db 124 AlaGlyThrArgSerAsnSerValGlyTyrAspArgPheIleAlaGluHisLeuAspAla 143
Qy 1120 ACCAGCCACAACTTCTGGTCCAAG-----CCCTCTGGTACTTC 1158
Db 144 ThrThrLysAlaTyrTrpSerLysArgThrLeuSerGlyArgArgHisSerValPhe 163
Qy 1159 CAGCAGCGCTGTACTACAGGCGGATGGGCAAGCTGTGGTGGTGTGCTGCAGTGGCTG 1218
Db 164 AspArgAsnIleTyrArgThrGlyLeuLeuGlyArgPheIleGlyAlaGlyHisIleMet 183
Qy 1219 GCCTGTGTGGTGGTGGGCAAGCGCTCAAGCGCTCGCAACGGCCGCCACATGGAG 1278
Db 184 AlaArgLeuHisGlyValLysLeuThr-----GluMetAlaLysThrArgThrLeuAsp 201
Qy 1279 GAGCAGCGCGCTGTGTGGGACAGCAATGCTCATCTTCATCTTCGTGTAAGAACGGGCCCAAG 1338
Db 202 GluGlnArgGlnPhePheAspSerLysValAlaProLeuPhe-----AspLys 217
Qy 1339 CCGCTGGTG---TGCTGTGTTCTCAAGTTCGTGAGCTGTGTCTTCAACAAGGCCGTG 1395
Db 218 ProValValArgTrpLeuThrLysArgLysSerSerLeu----- 230
Qy 1396 CTGTGTTGGCGCGGCTGCGGCGGAGCAGTAC-----GGCTGATCAAGCG 1446
Db 231 -----PheGlyLeuGlyIleProProArgGlnTyrAspGluLeuAlaSerLysSer 248
Qy 1447 GACGGCATCCCATTTAGAACTACATCGCGCGCACCATGACGCGGTGGCGGAACTCG 1506
Db 249 AspGly---ThrValAlaSerValLeuLysGluArgLeuGluLysLeuAlaCysAsnPhe 267
Qy 1507 CACGTGGCAAGCAACTACTTCTACTACAACCTGCCTCACCGGCAAGTTCCTCGCGAC 1566
Db 268 ProLeu---SerAspAsnTyrPheAlaTrpGlnAlaPheAlaArgTyrTrpGluPro 286
Qy 1567 AAC-----TGCCCACTACTCTCGCGAGGCGGCTTCGCCACCTCAAGATGGC 1617
Db 287 HisGluGlyAlaLeuProAlaTyrLeuLysProGluTyrTyrGluLysIleArgAsnAsn 306
Qy 1618 GTGGTGGACAACCTGACCGCTCTCCACCAACTTCTTCATGAGGAGCTCAAA----- 1668
Db 307 ThrAla---ArgValAlaValHisAlaThrTyrThrGluLeuLeuSerArgLysPro 325
Qy 1669 CGCGGCACCTACCAAGGTGATTCTGATGACCAACGCGGACTGCTGTGATATGCCCGTG 1728
Db 326 AlaAsnGlyValAspArgTyrIleLeuLeuAspAlaGlnAspTrpMetThrAspValGln 345
Qy 1729 CCCAACGAGCTGGCGAGTGTGGCCCAAGCAGGTTCGCGCGGCGGATCGTCACTGG 1788
Db 346 LeuAsnGluLeuTrpSerGlnIleSerArgThrAlaAlaSerGlyAlaArgValIlePhe 365
Qy 1789 CGCTCCGCTCC-----CTCAGCGCGCCCTACGCGGAGCTG 1824
Db 366 ArgThrAlaAlaGluLysSerValIleGluGlyArgLeuSerPro----- 380
Qy 1825 ATCCAGAGCGCGGCTTCAGCTGCGC-----TGCAATCCGCGCGCCACT 1869
Db 381 -----AspIleArgAsnGlnTrpValTyrLeuGluGluArgSerAsn 394
Qy 1870 CAG---GCTATACGACCGGTCAACATGTACAGCTCTCTTACATGGCCCGCGG 1923
Db 395 GluLeuAsnAlaMetAspArgSerAlaIleTyrGlyGlyPheHisIleTyrGlnArg 413
```

RESULT 2

```
AG2836
conserved hypothetical protein Atu2119 [imported] - Agrobacterium tumefaciens (strain C5)
C/Species: Agrobacterium tumefaciens
C/Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C/Accession: AG2836
R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, E
ster, E.W.
A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A/Reference number: AB2577; MUID:21608550; PMID:11743193
A/Accession: AG2836
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-416 <KUR>
A/Cross-references: UNIPROT:Q8UDK6; UNIPARC:UPI000000D1DB1; GB:AB008688; PIDN:AAL43109.1
A/Experimental source: strain C58 (Dupont)
C/Genetics:
A/Gene: Atu2119
A/Map position: circular chromosome
Alignment Scores:
Pred. No.: 4,59e-11 Length: 416
Score: 326.50 Matches: 111
Percent Similarity: 45.1% Conservative: 69
Best Local Similarity: 27.8% Mismatches: 158
Query Match: 9.0% Indels: 61
DB: 2 Gaps: 16
US-10-620-914-44 (1-1947) x AG2836 (1-416)
```

```
Qy 823 CTGTACAGCAGCTGTGGGAGGACCCGCGGATATGAGGTGATGGATCAACCCC 882
Db 44 ValTyrProGlnIleTrpGluAspProGluIleAspMetGluAlaMetGluLeuGlyGlu 63
Qy 883 AAGGACACGCTGTGACCCCTGACTAGCGCGGCTGCAATGCCCTGAACCTGCTGTGAG 942
Db 64 GlyHisArgIleValThrIleGlySerGlyCysAsnMetLeuAlaTyrLeuSerArg 83
Qy 943 GGGCGCGCCAGGTGTGTGCTGGTGCAGTCAACCCCGCAGTGGCGCTTCTGTGAGCTG 1002
Db 84 AsnProAlaSerIleAspValValAspLeuAsnProHisIleAlaLeuAsnLysLeu 103
Qy 1003 AAGAGTGTGCCATTACAGAGCTG---GAGTTTGGAGACGTGTGGAGCTTCTCGCGAG 1059
Db 104 LysLeuAlaAlaPheArgHisLeuProAlaHisGlnAspValValArgHisPheLysArg 123
Qy 1060 GGCGTGACCCGCCATTGAGGAGCTGTACAGAAAGAGCTGGCGCCCTTCTCTGTCCAA 1119
Db 124 AlaGlyThrArgSerAsnSerValGlyTyrAspArgPheIleAlaGluHisLeuAspAla 143
Qy 1120 ACCAGCCACAACTTCTGGTCCAAG-----CGCTCTGGTACTTC 1158
Db 144 ThrThrLysAlaTyrTrpSerLysArgThrLeuSerGlyArgArgHisSerValPhe 163
Qy 1159 CAGCAGCGCTGTACTACAGGCGGATGGGCAAGCTGTGGTGGTGTGCTGGTGGCTG 1218
Db 164 AspArgAsnIleTyrArgThrGlyLeuLeuGlyArgPheIleGlyAlaGlyHisIleMet 183
Qy 1219 GCCTGTGTGGTGGGCAAGCGCTCAAGCGCTCGCAACGGCCGCCACATGGAG 1278
Db 184 AlaArgLeuHisGlyValLysLeuThr-----GluMetAlaLysThrArgThrLeuAsp 201
Qy 1279 GAGCAGCGCGCTGTGTGGGACAGCAATGCTCATCTTCATCTTCGTGTAAGAACGGGCCCAAG 1338
Db 202 GluGlnArgGlnPhePheAspSerLysValAlaProLeuPhe-----AspLys 217
Qy 1339 CCGCTGGTG---TGCTGTGTTCTCAAGTTCGTGAGCTGTGTCTTCAACAAGGCCGTG 1395
Db 218 ProValValArgTrpLeuThrLysArgLysSerSerLeu----- 230
```



Db	440	GlyAlaProGlyThrProAlaAlaProGlyPro-----GlyGlyGlyAla	454
Qy	892	CCGTTCTCTGGGTTGATCTCCA-----TCACCTCCA	860
Db	455	AlaValProSerGlyAlaThrProHisProGluArgGlySerGlyProAlaAspProPro	474
Qy	859	TATCCGGCTCGGGTCTCTCCAAGCATGCGTGACAGGAAGTGGCGGGAACATGGGCG	800
Db	475	AlaAlaAlaArgLeuProProGluArgGln---GluProArgLeuProGlnAspLeuAla	493
Qy	799	CGCGCTCCAGCGCTCTCTGTCAGGCGTGGCCACGCTGGCGAGCGGCCCAATCCACA	740
Db	494	AlaAlaGlnArgCysPro-----AlaGlyProProProThr	505
Qy	739	CGTAGTAGGGCGCGCAGCACGGCAGTAGGGATCGAAACCTGGGTCTCTGCTCCC	680
Db	506	ArgSer---GlyAlaAlaGlnArgThrHisArgArgProProGlyCysProArgSer	524
Qy	679	ACACGCGCTCCAGCTTCTGCTCCAGGTAGCGCGCGCTCGGGCGCGATGTCAAATGTTG	620
Db	525	AlaArgAsnProGly-----	529
Qy	619	CGATGTCGAAGATCGATCGCCAGAAGAACGGCGGACACAGGCA-----TCTGCG	569
Db	530	---CysProArgThrTrpArgArgSerGlyAlaGlnArgGlyHisProProGly	548
Qy	568	GCAGGGCGAGTCTGTACTTCCGCTCACGTAGAAGTCGGCAGCGCCACAGAGCCGCTT	509
Db	549	AlaGly-----GlnArgProSerGlyProThr	557
Qy	508	GGGACAGGTACGAAACCTGTCGATGACGTTGTGGAACGCTGGAATCATCGTCAGCG	449
Db	558	GlyGlyArgProAlaAlaProGly-----	565
Qy	448	AGTAGGAGAGGTGATGAGGTGCGGTGCGCTCAGGGGGCGCAATGGCAAGCTCGG	389
Db	566	-----AlaPro-----GlyThrProAla	571
Qy	388	CCTCCAGCACTGGACATTTCCAGCCCTTGGCCTTCGCTTCTTGGCCACCTCGC	329
Db	572	AlaProGlyProGlyGlyAlaAlaVal---ProSerGlyAlaThrProHisProGlu	590
Qy	328	ACAGCAGTGGCACAGGTGCAACCACTAGTAGGACTTGAATTCGCGCAGGTGATGTAAT	269
Db	591	ArgGlySerGlyProAlaAspPro-----ProAlaAlaAlaArg	603
Qy	268	CAGCCATCATTCACATTTCTCCCATGTCGCCACCCAGGTCAACCCAGATGAGTTTCG	209
Db	604	LeuPro-----ProGluArgGlnGluProArgLeuProGlnAspLeuAla	618
Qy	208	AGCCCTCGG-----CCAGGGGCGCAGAACGCGGCTGGGGCCGTAGAGTCT	155
Db	619	AlaAlaGlnArgCysProAlaGly-----ProProProThrArgSerGly	633
Qy	154	CCAGGCGCAGCGTGTATCATCGC-----CCTTCTTGTGCGGAACCAATATGGCGCA	101
Db	634	AlaAlaAlaGlnArgThrHisArgArgProProGlyCysProArgSer-----	649
Qy	100	GAACGGTCAGGTCACTCTTCATGC	77
Db	650	-----AlaArgAsnProGlyCys	655

## RESULT 4

hypotheical 119.5K protein (uvra region) - Micrococcus luteus  
 N:Alternate names: ORF 1 protein  
 C:Species: Micrococcus luteus, Micrococcus lysodeikticus  
 C:Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 04-Feb-2000  
 C:Accession: JQ0405  
 F:Shiota, S.; Nakayama, H.  
 Mol. Gen. Genet. 217, 332-340, 1989  
 A:Title: Micrococcus luteus homolog of the Escherichia coli uvra gene: identification of  
 A:Reference number: S04781: MUID: 893164717: PMID:2549377











```
QY 1523 ACTACTTACTCAACTGCTCACC- - - - -GCAAGTTCTCGCGCACTGCCCCA 1576
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 390 AlaThrAlaAlaThrAlaAlaThrProAlaArgAlaAlaThrProAlaThr 409
QY 1577 CTTACTCGCGAGCGCGCTTCCGACCTCAAGAGTGGCGTGGACAACTGACCG 1636
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 410 ProAlaThrProAlaThrProAlaThrProAlaThrAlaAlaThrAlaAla 429
QY 1637 TCTCCACCAACTTCTTCATGG- - - - -AGGAGCTCAAAGCGCGCACTACCA 1684
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 430 ThrAlaAlaThrAlaAlaThrAlaAlaThrAlaAlaThrAlaProThr- - - 448
QY 1685 AGGTGATCTGATGGACCGTGGAGTGGCTGGATATGCCGCGGCAAGAGTGGCCG 1744
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 448 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - 448
QY 1745 AGTGCCTGGCAAGCAGTTGGCGCGGGCGCATCGTCATCTGGCGCTCGCCCTCCCTCA 1804
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 449 - - - - -ProAlaArgAlaAlaThrAlaAlaThrProAlaThrGlyAlaThrProAla 465
QY 1805 GCCCGCCCTACGCGGAGCTGATCCAGAAAGCGGGCTTCGACGTGGCTGCATCCGCGCG 1864
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 466 ThrAlaProThrAlaGly- - -ThrAlaAlaThrAlaAlaThrAlaAlaThrAlaAlaThr 484
QY 1865 CCACTCAGGCTACATGACCGCGTCAACATGTACAGCTCTCTTCTACA 1912
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 485 ProAlaArgAlaSer- - -ThrProAlaThrAlaAlaThrProValThr 499

RESULT 8
S27923
gene LP3 protein - human herpesvirus 4
C:Species: human herpesvirus 4, Epstein-Barr virus
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 09-Jul-2004
C:Accession: S27923
R:Parker, B.D.; Bankier, A.; Satchwell, S.; Barrell, B.; Farrell, P.J.
submitted to the EMBL Data Library, August 1990
A:Description: Sequence and transcription of Raji Epstein-Barr virus DNA spanning the B9
A:Reference number: S27923
A:Accession: S27923
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-924 <PAR>
A:Cross-references: UNIPROT:Q99307; UNIPARC:UPI00000615C0; EMBL:M35547; NID:g330420; PID
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;

Alignment Scores:
Pred. No.: 1,31e-07 Length: 924
Score: 267.00 Matches: 196
Percent Similarity: 30.8% Conservative: 40
Best Local Similarity: 25.6% Mismatches: 296
Query Match: 7.4% Indels: 234
DB: 2 Gaps: 41

US-10-620-914-44 (1-1947) x S27923 (1-924)
QY 1888 CGCGGTCCATGT- - - - -AGCCCTCAGTGGCGCGCGGA 1856
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 6 ArgGlyProCysLeuAlaProGlySerGlyLeuGlyAlaHisProHisProArgArgSer 25
QY 1855 TGCAGCGCAGCTGCAAGCCGCGCTTCTGGA- - - - -TCAGTCCGGGTAGGGCGGCTGA 1802
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 26 GlyAlaAlaAspProAlaAspProValGlyHisProAlaAlaProArgAlaProGly- - - 44
QY 1801 GGGAGCGGAGCGCCAGATGACATCGCCCGCGCACTCTCTGGCCAGGCACTCGG 1742
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 45 - - - - -ProGluProArgThrArgLeuGlnProAlaThrProArgArgSerGlyAlaAla 62
QY 1741 CCAGCTCGTGGCGCAGGCATATCCAGCCAGTCCACGTGGTCCATCAGAAATCACCTTGG 1682
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 63 AspProAlaAspProValGlyHisProAlaAlaProArgAlaProGlyProGluProArg 82
QY 1681 TGTAGTGGCGCGCTTTGAGCTCTCTCCATGAAGAAGTTGGTGGAGACGGTCAGGTTGT- - - 1625
```

```
Db 83 ThrArgLeuGlnProAlaThrPro- - - - -ArgArgSerGlyAlaAla 96
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 1624 - - - - -CCACACCGCACTCTTGAGGTGGCGAAGCG- - - - -CCTCGC 1586
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 97 AspProAlaAspProValGlyHisProAlaAlaProArgAlaProGlyProGluProArg 116
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 1585 GCAGGTAGTGGGGCAGTTGTGCGCAGAACTTGCCTGGTGGAGCAGTGTGTAGTAGAAGT 1526
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 117 Thr- - - - -ArgLeuGlnProAlaThrProArg 125
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 1525 AGTTCTGCTTGGCAGCTGGAGTTCTCCGACAGCGCTTCATGTGGCGCGCATGTAGT 1466
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 126 ArgSerGlyAlaAlaAspProAlaAspProValGlyHisProAlaAlaProArgAlaPro 145
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 1465 TCTCAATGCGGATGCCCTCGCTTGCATCAGCGCTACTGCTTGGCGCGCACCGCCGCGC 1406
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 146 - - - - -GlyProGluProArgThrArgLeuGlnProAlaThrPro- - - - -ArgArg 160
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 1405 CGAACACACAGCAGCGCT- - - - -TGTGAAGAGCAGCACAGC 1370
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 161 SerGlyAlaAlaAspProAlaAspProValGlyHisProAlaAlaProArgAlaProGly 180
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 1369 TCAGCACTTGAGCAACACACACAGCGGCTTGGCGCGCTTCTTCACGAAGTGGATGA 1310
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 181 ProGluProArgThrArgLeuGlnProAlaThrProArgArgSerGlyAlaAlaAsp- - - 199
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 1309 GCATGTTGCTGCCACAGCGCGCTGCTCTCCATTTGGCGCGCTTGGCGAGCGCT 1250
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 200 - - - - -ProAlaAspProValGlyHisProAlaAlaProArg- - - - - 211
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 1249 TGACGGCTTGTCCAGTCCAGCACACAGC- - - - -CCAGCACTGACGACACC 1202
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 212 - - - - -AlaProGlyProGluProArgThrArgLeuGlnProAlaThrProArgArg 228
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 1201 AGCAGAGCTTG- - - - -CCATCGCGCTTGTGTAGTACAGC 1166
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 229 SerGlyAlaAlaAspProAlaAspProValGlyHisProAlaAlaProArgAlaProGly 248
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 1165 CGTGTGGAAGTACCAGA- - - - -GGCGCTTGGACCAAGTTGTGGCTTTCGACCA 1112
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 249 ProGluProArgThrArgLeuGlnProAlaThrProArgArgSerGly- - -AlaAlaAsp 267
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 1111 GGAAGGCGCCAGCTTCTTCGTACAGCTCTCAATGCGCGGT- - - - - 1067
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 268 ProAlaAspProValGlyHisProAlaAlaProArgAlaProGlyProGluProArgThr 287
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 1066 - - - - -GCAGCGCTTCCGCAACAGTCCACAGCTCTCAACTCCAGCTGCT 1019
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 288 ArgLeuGlnProAlaThrProArgArgSerGlyAlaAlaAspProAlaAspProValGly 307
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 1018 GAATGCGCACCTTCTTCAGCTCCAGAAAGCGCGAGCTGCGGGGTTCAGTCCACCGACA 959
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 308 His- - - - -ProAlaAlaProArgAlaProGlyProGluProArgThr 321
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 958 - - - - -CCACTGGC 950
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 322 ArgLeuGlnProAlaThrProArgArgSerGlyAlaAlaAspProAlaAspProValGly 341
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 949 CGGCCCCCTGACACAGCAGGTTTCAGGGCATTCGACGCGC- - - - - 911
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 342 HisProAlaAlaProArgAlaProGlyProGluProArgThrArgLeuGlnProAlaThr 361
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 910 - - -CGTACTCAGGTTCAGCA- - -CGTGT- - -CCTTGGGTGTGCTCCATCAGCTCCA 860
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 362 ProArgArgSerGlyAlaAlaAspProAlaAspProValGlyHisProAlaAlaProArg 381
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 859 TATCCGGCTCGGGGTCTCCACG- - - - -ACTCGGTGTACAGAAAGTGGCGGGAACA 806
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 382 AlaProGlyProGluProArgThrArgLeuGlnProAlaThrProArgArgSerGlyAla 401
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 805 - - - - -TGGCGCGCGCTCCAGCGCTCTCTGTGAGGCGCT 770
    |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
```

402 AlaAspProAlaAspProValGlyHisProAlaAlaProArgAlaPro----- 417  
QY 769 GGCCAAAGCTGGGAGCGGCCAATCCACAGTAGTAGGGGGCGCGCAGCCAGCGACGT 710  
Db 418 -----GlyProGluProArgThrArgLeuGlnProAlaThrProArg 431  
QY 709 AGGGATCGAACCTGGGTGTTGCTCCACACGCGCTCCAGCTTCCTCCAGGTAGG 650  
Db 432 ArgSerGlyAlaAlaAspProAlaAspProValGlyHisProAlaAlaProArgAla 450  
QY 649 CGCGCGCTCGGGCCGATGTCATGTGTGTCGATGTCGAAGATCGATCGCCAGAGAAC 590  
Db 451 ProGlyProGluProArg-----ThrArgLeuGlnProAlaThrProArgArgSer 467  
QY 589 GGCGCGACCGAGCATCTGGCGAGGGGAGGTGCTACTTGGCGCTCAGTAGAGTCGG 530  
Db 468 GlyAlaAlaAspProAlaAspPro-----ValGlyHisProAlaAlaProArgAlaProGly 486  
QY 529 CAACGCCACCGAGCGCTCTGGACAGGTACGAGCAAGCTGCTCGATGAGTGTGGA 470  
Db 487 ProGluProArgThrArgLeuGlnProAlaThrProArg-----ArgSerGly 502  
QY 469 ACGTGTGAATCATGTCGAGCGAGTAGGAGAGGTGATGAGCGTCGCGGTCCCTCAGGG 410  
Db 503 AlaAlaAspProAla-----AspProValGly 511  
QY 409 GCGCAATTCGCAAGCTCGGCT-----CCAGACCT----- 377  
Db 512 HisProAlaAlaProArgAlaProGlyProGluProArgThrArgLeuGlnProAlaThr 531  
QY 376 -----GAGCATCTTCAGCCCT--- 359  
Db 532 ProArgArgSerGlyAlaAlaAspProAlaAspProValGlyHisProAlaAlaProArg 551  
QY 358 TGGCTTCGCTTCCTCTGGCCACCTCGCACAGCGAGTGGCACAGGTTCGACCAAGCTAGA 299  
Db 552 AlaProGlyProGlu-----ProArgThrArgLeuGlnProAlaThrProArgArg 568  
QY 298 TGGACTTGAATTCGCGAGTCGATGTAATGATGATGATGATGATGATGATGATGATGATG 239  
Db 569 SerGlyAlaAlaAspPro-----AlaAspProVal 578  
QY 238 CACACCCAGGTCAACCCAGATGAGTTTCGAGCGCTCGGCGAGCGGCGACCAAGCAG 179  
Db 579 GlyHisProAlaAlaProArg-----AlaProGlyProGluProArgThr 593  
QY 178 CGG----- 176  
Db 594 ArgLeuGlnProAlaThrProArgArgSerGlyAlaAlaAspProAlaAspProValGly 613  
QY 175 ---CCTGGGGCCGTAGAGCTCTCCAGGCGAGCGGTGATCAT---CGCCCTTCTTGC 122  
Db 614 HisProAlaAlaProArgAlaProGlyProGluProArgThrArgLeuGlnProAlaThr 633  
QY 121 TCCGGAACCATATGCGCGAGACGCTCA-----GTCATCTTCATGCTGCTGA 71  
Db 634 ProArgArgSerGlyAlaAlaAspProAlaAspProValGlyHisProAla----- 650  
QY 70 GCTTGAGCTTCTCCAGGG-----AGAAAGTTCTTCTTGGTGTAGCTCGCAG 26  
Db 651 AlaProArgAlaProGlyProGluProArgThrArgLeuGlnProAlaThrProArgArg 670  
QY 25 GCGGCGCTCAGACCG 8  
Db 671 SerGlyAlaAlaAspPro 676

## RESULT 9

F75518  
hypothetical protein - Deinococcus radiodurans (strain R1)  
C;Species: Deinococcus radiodurans  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C;Accession: F75518  
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;

M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Mal  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A;Reference number: A75250; MUID:20036896; PMID:10567266  
A;Accession: F75518  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-839 <WHI>  
A;Cross-references: UNIPROT:Q9RX57; UNIPARC:UPI00000C1761; GB:AE001904; GB:AE000513; NID  
A;Experimental source: strain R1  
C;Genetics:  
A;Gene: DR0458  
A;Map position: 1  
A;Superfamily: herpesvirus immediate-early protein IE110; RING finger homology

Alignment Scores:  
Pred. No.: 1.68e-07 Length: 839  
Score: 265.00 Matches: 157  
Percent Similarity: 31.8% Conservative: 51  
Best Local Similarity: 24.0% Mismatches: 242  
Query Match: 7.3% Indels: 204  
DB: 2 Gaps: 27

US-10-620-914-44 (1-1947) x F75518 (1-839)

QY 1771 CCGCGCAACCTCTTGGCCA-----GGCACTCGGCAGCTCGTGGCCACGGGCA 1721  
Db 143 ProAlaGluProLeuLysProProValGlnAspThrProProValThrProLysPro 162  
QY 1720 TATCCAGCCAGTCACGTGTCATCAGATCAATCACTTGTGTAGTGTGGCGCTTGGAGCT 1661  
Db 163 ValThrProGluProValThrProLysProAlaPro-----Thr 175  
QY 1660 CTTCCATGAAGATTGTTGGACAGCGTCAAGTTGTCACACCGCCACTCTTGGAGGTGG 1601  
Db 176 ProGluValLeuGlnProProValAlaGlnThrProPro-----Val 190  
QY 1600 CGAAGCGCGCTCGCGCAGTAGTGGGCGAGTTGTTCGCCAGGAACCTTCCCGTGAGGC 1541  
Db 191 AlaLysProPro----- 194  
QY 1540 AGTTGTAGTAGAAGTAGTCTCTGTCGCAAGTCGAGTTCTCGCCACCGCTCCATGG 1481  
Db 195 -----ValProAlaProThrSerGlnThrProThrProValGlnProAla 210  
QY 1480 TGGCGCGAGTAGTCTCAATGGGATGCGCTCGCTTGTATCA-----GCGGTACT 1427  
Db 211 ProThrArgThrProProGlnAlaAlaArgProThrProAsnAlaProAlaGlnThr 230  
QY 1426 GCTTGGCCGCGCAGCGCCGCCGACCCACAGCAGCGGCTTGTGAAGAGCAGCAGGCTCA 1367  
Db 231 ProAlaProAlaThrGlnAlaProAlaAlaGlnThrProThrAlaGlnAlaProAlaThr 250  
QY 1366 CGAATCTTGACGAACAGCCACACAGCGGCTTGGGCCCGCTTCTTCACGAAGTGGATGAGCA 1307  
Db 251 GlnThrProAlaThrProAlaProAlaAlaGlnArg----- 262  
QY 1306 TGTGTGTCTCCACAGCGCGCTGCTCTCCATGTGTGGCGCGCTTGGCGAGGCGCTTGA 1247  
Db 263 -----ProAlaGlyGlyAla----- 267  
QY 1246 CGGTCTTCCAGTCCCGCAGCAGCCAGCGCCAGGCACTGCA----- 1208  
Db 268 -----ProSerProAlaProAlaProAlaGlnAlaAsnAlaProAlaGlySerVal 284  
QY 1207 -----GCACCCAGCAGCACTTGGCCCATCGCGCTGTGTACAGGCGCGT 1163  
Db 285 ValProGluAlaThrValProGluSerSerThrProAlaAlaProSerAlaGlnThrPro 304  
QY 1162 GCTGGAAGTACAGAGCGCTTGGACCAAGAGTTGTGGCTGTGGACAGGAGGGCG 1103  
Db 305 ProThrProThrArg-----GluThrAlaGlnThrGluAlaSerProAlaAla 320

1102 CCAGCTCTTCTCAGCTCCTCAATGCGCGGTGCACGCCCTCGCCGACAGCTGCC 1043  
||| ||||| ||||| : : |||  
321 ProAsnSerSerAlaAlaProAsnGluProAlaSerGluPro-----ValAla 337  
1042 ACAGTCTCTCAACTCAGCTGCTGAATGCGCCACTTCTTCAGCTCCAGAACGCCGACT 983  
||| ||||| ||||| ||| : : : : : |||  
338 GlyArgProGlyThr---AlaAlaSerSerProGluSerAlaSerProValThrValThr 356  
982 GCCGCGGTTGCACTCCACCGACA-----CCACTGCGCGCGGCC----- 944  
||| ||||| : : ||| ||| ||| |||  
357 ProArgGlyGluThrProAspThrAlaAlaSerAlaGlyThrProSerAlaGlyArgVal 376  
943 ---CCTGCACACAGCTTCAGGCAATGTCAGCGCGCGCTAGTCAGGGTCAGACCGTGT 887  
||| ||||| ||||| ||| : : : : : |||  
377 ThrProAlaProAlaProSer-----AlaSerGluGlyAlaSerAlaAlaArgThr 393  
886 CTTTGGGTGATCTCCATCAGCTCCATATCCGGCTCGGGTCTCTCCAGCACTGGTGT 827  
||| ||||| ||||| ||||| ||||| |||||  
394 ProGlyAlaGlySerGlnThrProProleProAlaThrProleProAlaThrProAla 413  
826 -----ACAGAGAGTGGCGGGAACATGGCGCGCTCCACGCGCTCCTCGTGCA 776  
||| ||||| ||||| ||||| ||||| |||||  
414 GlyArgSerSerGlyGluSerAlaGlyThrAlaAlaAlaArgProAlaAlaProAla 433  
775 -----GGCGGTGGCCCAACGCTGG---GCAGCGCGC 749  
||| ||||| ||||| ||||| ||||| |||||  
434 ProValSerGluAspArgSerAspValSerGlyLeuProArgArgGluAspAlaProAla 453  
748 CAATCCACA-----CCTAGTGGGCGCGCAGCCAGCCAGCTAGGATCGAAC 698  
||| ||||| ||||| ||||| ||||| |||||  
454 GluSerSerProValAlaAlaSerProAlaArgGlyAlaSerSerAlaProSerSerAla 473  
697 CTTGGGTGTTCTGCTCCACACGCGCTCCAGCTTCT----- 662  
||| ||||| ||||| ||||| ||||| |||||  
474 ProAlaAlaAlaValProSerArgAlaProValSerGlySerValSerAlaProArg 493  
661 -----GCTCCAGTTCAGCGC 647  
||| ||||| ||||| ||||| ||||| |||||  
494 ThrAlaProThrAlaProValAlaGluGlnGlyGluValProValSerProSerAlaAla 513  
646 GCGCTCGGGCGCGTCAATGTTGCGATGTCGAGATCG---ATCGCCAGAACAAAC 590  
||| ||||| ||||| ||||| ||||| |||||  
514 AlaProArgGlyGlyAlaSerSerAlaAlaAlaProSerAlaProAlaAlaArgGly 533  
589 GCGCGCACACGCGCATCTGCGCAGCGGCGAGTCTGCTTTCGCTCAGCTAGAGTCG 530  
||| ||||| ||||| ||||| ||||| |||||  
534 GlySerGlyAlaAlaGlyAlaAlaGlyGlyAlaSerAlaProAlaAlaAlaArgPro 553  
529 CAACGCCACCGCGCTTTGGGACAGGTAACGCAAGCTGTCGATGACGTTGTGGA 470  
||| ||||| ||||| ||||| ||||| |||||  
554 AlaGlnThrProGlyAlaSerAlaGlyAlaSerGlyGlyGly----- 568  
469 ACGTGTGATCATCTGTCAGCGATGAGAGAGTGTATGCGTTCGCGTCCCTCAGGG 410  
||| ||||| ||||| ||||| ||||| |||||  
569 -----GluGly 570  
409 GCGCAATTTGGCAAGCTCGCGCTCCACGACCTGGACATTTCTTCCAGCGCTTCGCTCG 350  
||| ||||| ||||| ||||| ||||| |||||  
571 ValSer-----AlaArgProSerGlnGlyThrProSerGlyThr 584  
349 CTTTCTTCTTGGCCA-----CCTCGCACAGGAGTGGCACAGTTCGACCGCTAGA 299  
||| ||||| ||||| ||||| ||||| |||||  
585 ProAlaSerAlaProValAlaAlaGlyArgProAlaGlyGlyGly----- 600  
298 TGGACTTGAATCTGCGCAGTGCATGTAATCAGCCATCATATCGCATCTCCCGAGTGC 239  
||| ||||| ||||| ||||| ||||| |||||  
601 -----SerGlySerGlyThrSerGlySerGlyAlaProAlaAla 615  
238 CACCACCCAGGTCAA----- 221  
||| ||||| ||||| ||||| ||||| |||||  
616 AlaArgProGlyGlnGlyGlySerGlyAspAlaProValAspLeuGlyLeuThrSerGly 635

220 AGATGAGTTTCAGACGCTCGCGCGCGG----- 191  
||| ||||| ||||| ||||| ||||| |||||  
636 ArgGlyGlySerGlyAlaGlyAsnGlyGlyAlaGlyThrGlyGluThrGlyAlaGlyThr 655  
190 CAGCAAAAGCAGCGCGCTCGGGCC-----CGTAGAGCTCTCCAGCGGAGCAG 143  
656 ProSerAlaSerArgProGlySerGlyGlyThrGlyAsnGlySerGlyThrGlyGlyGly 675  
142 CGTGATCATCGCTCTTCTGCTGCCGAACACCATATGCGCA 101  
||| ||||| ||||| ||||| ||||| |||||  
676 SerGlyThrValProAlaSerGlyArgGlyThrGlyGlySer 689  
RESULT 10  
S48478  
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomyces cerevisiae)  
N;Alternate names: extracellular glucoamylase; mucin-like protein MUC1; protein; YIR019C  
C;Species: Saccharomyces cerevisiae  
C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004  
C;Accession: S48478; A26877; B26877; S27281; JG6123  
R;Rowley, K.  
submitted to the EMBL Data Library, October 1994  
A;Reference number: S48478  
A;Accession: S48478  
A;Molecule type: DNA  
A;Residues: 1-1367 <ROW>  
A;Cross-references: UNIPROT:P08640; UNIPARC:UPI000000056E; GB:Z47047; EMBL:Z38061; NID:946  
R;Yamashita, I.; Nakamura, M.; Fukui, S.  
J. Bacteriol. 169, 2142-2149, 1987  
A;Title: Gene fusion is a possible mechanism underlying the evolution of STA1.  
A;Reference number: A91831; MUID:87194600; PMID:3106330  
A;Accession: A26877  
A;Molecule type: DNA  
A;Residues: 1-242 <YAM>  
A;Cross-references: UNIPARC:UPI0000168E08; EMBL:M16164; NID:g172522; PIDN:AAA35014.1; PII  
A;Accession: B26877  
A;Molecule type: DNA  
A;Residues: 762-1331 <YAZ>  
A;Cross-references: UNIPARC:UPI0000168E09; EMBL:M16165; NID:g172523; PIDN:AAA35015.1; PII  
R;Pardo, J.M.; Tanex, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.  
FEBS Lett. 239, 179-184, 1988  
A;Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Sacchar  
A;Reference number: S27281; MUID:89031230; PMID:3141213  
A;Accession: S27281  
A;Molecule type: DNA  
A;Residues: 1-31 <PAR>  
A;Cross-references: UNIPARC:UPI0000168E60; EMBL:X13857; NID:g4551; PIDN:CAA32069.1; PID:5  
R;Lambrechts, M.G.; Bauer, F.F.; Marmur, J.; Pretorius, I.S.  
Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996  
A;Title: Muc1, a mucin-like protein that is regulated by Msa10, is critical for pseudohy  
A;Reference number: JG6123; MUID:96323237; PMID:8710886  
A;Accession: JG6123  
A;Status: nucleic acid sequence not shown  
A;Molecule type: DNA  
A;Residues: 1-1367 <LAM>  
A;Cross-references: UNIPARC:UPI000000056E; GB:U30626; NID:g1304386; PIDN:AAC49609.1; PID  
C;Genetics:  
A;Gene: SGD:MUC1; STA2; MAL5; DEX2; SGD:S0001458  
A;Cross-references: MIPS:YIR019c; SGD:S0001458  
A;Map position: 9R  
C;Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase  
C;Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein  
F;5-21/Domain: transmembrane #status predicted <TM1>  
F;1350-1366/Domain: transmembrane #status predicted <TM2>  
Alignment Scores:  
Pred. No.: 5.26e-07 Length: 1367  
Score: 257.00 Matches: 170  
Percent Similarity: 38.3% Conservative: 85  
Best Local Similarity: 25.6% Mismatches: 298  
Query Match: 7.1% Indels: 112  
Gaps: 24  
DB: 1  
US-10-620-914-44 (1-1947) x S48478 (1-1367)









Db 3400 ProGluThrThrHisThrSerThrValLeuThrThrAlaThrMetThrArgAlaThr 3419  
QY 1781 TCATCTGGCGCTCGGCTCCTCAGCC-----CGC 1810  
Db 3420 AsnSerThrAlaThrProSerSerThrLeuGlyThrThrArgIleLeuThrGluLeuThr 3439  
QY 1811 CCTACGCGGAGCTGATCCAGAGCGGGCTTCACGCTGCGCTGCATCCGCGCGCACTC 1870  
Db 3440 ThrThrAlaThrThrAlaAlaThrGlySerThrAlaThrLeuSerSerThrPro--- 3458  
QY 1871 AGGCGTACATGG-----ACGCG 1888  
Db 3459 GlyThrThrTrpIleLeuThrGluProSerThrIleAlaThrValMetValProThrGly 3478  
QY 1889 TCAACATGTACAGCTCTTACATGCGCGCGGAGGCGGCAAGAGGCAACT 1945  
Db 3479 SerThrAlaThrThrSerSerThrLeuGlyThrAlaHisThrProLysValThr 3497

RESULT 12  
T45463  
membrane glycoprotein [imported] - equine herpesvirus 1  
C;Species: equine herpesvirus 1  
C;Date: 31-Jan-2000 #sequence\_revision 31-Jan-2000 #text\_change 09-Jul-2004  
C;Accession: T45463  
R;Kirisawa, R.; Kobayashi, T.; Kawakami, Y.; Iwai, H.  
J. Equine Sci. 7, 79-87, 1996  
A;Title: Nucleotide sequences of open reading frames 1, 24 and 71 of an attenuated equine herpesvirus 1 glycoprotein homolog  
A;Reference number: 222973  
A;Accession: T45463  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-867 <KIR>  
A;Cross-references: UNIPROT:O39782; UNIPARC:UPI000000EF109; EMBL:D88734; PIDN:BAR20038.1  
A;Experimental source: isolate 3F clone; strain BK343  
C;Genetics:  
A;Note: ORF71  
C;Superfamily: equine herpesvirus glycoprotein X; equine herpesvirus 1 glycoprotein homolog

Alignment Scores:  
Pred. No.: 1.06e-06 Length: 867  
Score: 251.00 Matches: 119  
Percent Similarity: 41.0% Conservative: 81  
Best Local Similarity: 24.4% Mismatches: 208  
Query Match: 6.9% Indels: 80  
DB: 2 Gaps: 14

US-10-620-914-44 (1-1947) x T45463 (1-867)

QY 473 ACAACGTCATCGACCGCTTGTCTGCTCTCCCAAGCGGCTGGTGGCGTTCGCG 532  
Db 46 ThrThrAsnSerSerSerProThrThrSerProThrThrSerSerProPro 65  
QY 533 ACTTCTACGTGAGCGGCAAGTACGACCTGCCCTCGCCAGATGCCCTGTGCGCGCTT 592  
Db 66 ThrSerThrHisThrSerProSerThrSerThrGlnSerSerThrAlaAla 85  
QY 593 TCTTCTGGGATCATCTTGACATCGACACATTGATCGCGCGCGCGCGCT 652  
Db 86 ThrSerSerAlaProSerThrAlaSerSerThrSerIleProThrSerThr 105  
QY 653 ACCTGGAGCAGAACCTGGAGCGGTGGGAGCAGAACACCGAGGTTCGATCCCTACG 712  
Db 106 ThrGluThrThrThrThrProThrAlaSerThrThr-----ThrProThr 121  
QY 713 TGCGGTGGCTGCGCGCCCTTACTGATGTGGATTTGGCGCGCTGCCAGCGTGGCCACG 772  
Db 122 ThrThrThrAlaAlaProThrThrAlaAlaThrThrThrAlaValThrAlaAlaSer 141  
QY 773 CCCTGACGAGGCGGTGGAGCGCGCCATGTTCCCGCCCACTCTCTGTACAGCG 832  
Db 142 ThrSerAlaGluThrThrThrAlaThrAlaThrAlaThrProThrThr--- 160  
QY 833 AGTCGTGGAGGACCCCGAGCGGATATGGAGGTATGGAGATCAACCCCAAGGACACGG 892

Db 161 -----ThrProThr-----SerThrThrThr----- 168  
QY 893 TGCTGACCTGACTAGCGCGGCTCAATGCCCTGAACCTGCTGTGTCAGGGGCGCGCC 952  
Db 169 -----ThrAlaThrThrValProThrThr----- 177  
QY 953 AGGTGCTGTGGTGGACTGCAACCCCGCAGTCGGCGCTTCTGGAGCTGAAGAAGTGG 1012  
Db 177 ----- 177  
QY 1013 CCATTGACAGCTGGAGTTTGAGGACGTGTGGCAGCTGTTGGCGAGGGCGTGCACCCGC 1072  
Db 178 -----AlaSerThrThrThrAspThrThrAlaAlaThrThrAlaAlaThrThr 195  
QY 1073 GCATTGAGAGCTGTACGAGAGAACTGGCGCCCTTCCTGTGCGCAACCCAGCCACAAC 1132  
Db 196 ThrAlaAlaThrThrThrAlaAlaThrThrAlaAlaThrThrAlaAlaThrThr 215  
QY 1133 TCTGTCTCAAGCGCTCTGTACTTCCAGCAGCGCTGTACTACAGGCGCGCATGGCA 1192  
Db 216 ThrAlaAlaThrThrThrAlaAlaThrThrAlaAlaThrThrAlaAlaThrThr 235  
QY 1193 AGCTGTGCTGGTGTGTCAGTGCCTGGCGCTGTGTGGACTGGGCAAGCCGTCAAGC 1252  
Db 236 ThrAlaAlaThrThrThrAlaAla-----ThrThrThr 246  
QY 1253 GCCTGCCCAACCGCCCAACATGGAGGAGCGCGCTGTGTGGACAGCAACATGCTCA 1312  
Db 247 AlaAlaThrThrThrAlaAlaThrThrSerSerAlaThrThrAlaAlaThrThrAla 266  
QY 1313 TCCACTTCGGAAGAACCGGCCCAAGCGCTGTGTGGCTGTTCGTCACAGTTCGTGAGCC 1372  
Db 267 AlaThrThr-----ThrAlaAlaThrThrThrAlaAlaThrThrAlaAlaThrThr 284  
QY 1373 TGGTGTCTTTCAAGCGCTGTGTGTTCGGCGCGCGCTGTGGGCAAGCAGTACG 1432  
Db 285 ThrAlaAlaThrThrThrAlaAlaThrThrAlaAlaThrThrAlaAlaThrThr 304  
QY 1433 CGCTGATCAAGCGGAGCGCA-----TCCCATTTGAGACTACATCGCGCGCACCATGG 1486  
Db 305 ThrAlaAlaThrThrThrAlaAlaThrThrAlaAlaThrThrAlaAlaThrThr 324  
QY 1487 ACGCGTGGCGGAGAACTCGCAGCTGCGCAAGCAGAACTACTTCTACTACA----- 1537  
Db 325 ThrAlaAlaThrThrThrGlySerProThrSerGlySerThrSerThrGlyAlaSer 344  
QY 1538 ACTGCTCACCGCAAGTTCCTGCGGCAACTGCGCCCACTACCTCGCGGAGCGGCT 1597  
Db 345 ThrSerThrProSerAlaSerThrAlaThrSerAlaThrProthr----- 359  
QY 1598 TCGCCACCTCAAGATGGCGTGTGGACAACTGACCGCTCCCACT-----TCT 1651  
Db 360 SerThrSerThrSerAlaAlaAlaThrThrSerThrProThrProThrSerAlaAlaThr 379  
QY 1652 TCATGGAGGAGCTCAAGCGCGCACCTACACCAAGGTGATTCGATGGACACCGTGGACT 1711  
Db 380 SerAlaGluSerThrThrGluAlaProThrSer-----ThrProThrThr 394  
QY 1712 GGCTGGATATGCCGTGGCCCAAGAGCTGCGGAGTGGCGGAGTGGCCAGCAGGTTCGCGCG 1771  
Db 395 AspThrThrThrProSerGluAlaThrThrAlaThrThrSerProGluSerThrVal 414  
QY 1772 CGCGCATCTCATCTGGCGCTCGCCCTCCCTCAGCCCGCTTACCGCGAGCTGATCCAGA 1831  
Db 415 SerAlaSerThrThrSerAla-----ThrThrThrAlaPheThrThrGluSerHisThrSer 433  
QY 1832 AGCGCGCTTCGAGCTGCGCTGCATCGCGCGCGCACTCAGGCGCTACATGACGCGCTCA 1891  
Db 434 ProAspSerSerThrGlySerThrSerThrAlaGluProSerSerThrPheThrLeuThr 453  
QY 1892 ACATGTACAGCT---CCTTCTACA 1912

783	QY	GGAGCGCGTGAGCGGCC-----	800
761	Db	 GlyLeuGlyGlyLeuGlyAlaGluLeuLeuValGlyGlnArgValValGluHisHis	780
801	QY	-----GCCCATGTT-----       HisAlaHisValLeuGlyValGlyTyrLeuPro---HisProGlyGlyAlaAlaAlaGlu	851
781	Db	 HisAlaHisValLeuGlyValGlyTyrLeuPro---HisProGlyGlyAlaAlaAlaGlu	799
852	QY	CGCGATATATGGAGTGATGGAGATCAACCCCAAGGACACGGTGCTGACCTGACCTAGCGG	911
800	Db	 ArgGlyAlaAlaAlaArgGlyAspValArgGlnGlyGlyArgValGluGlyGluArgArg	819
912	QY	CGG----- -----CTGCATCGCCCTGAACCTGCTGCTGTCAGGGGGCCCGGCCA	953
820	Db	AlaProGluPheGlyGluAspLeuLeuValHisGluGlyAlaGlyHisLeuGlyArgAla	839
954	QY	---GGTGGTGTCGGTGGATG---       ValGlyGlyGluGlyArgLeuGlyGlyProArgArgValGlyLeuAlaGlyArgAsp---	1007
840	Db	 ValGlyGlyGluGlyArgLeuGlyGlyProArgArgValGlyLeuAlaGlyArgAsp---	858
1008	QY	GGTGGCCATTACGACGCTGGAGTTTCAGGACGCTGTCGACGCTCTTCGGCGAGCG---	1061
859	Db	----- -----AlaAlaGluAlaAlaValGlyArgGlyValLeu	869
1062	QY	-----CGTGCACCC 	1070
870	Db	GlyHisGlyProGluAlaAlaProGluProValValLeuGlyGlyGlyGlyGlyGly	889
1071	QY	-----GCCATTGAGGACCTGTACGA-----       GlnGlnArgGlySerGlyValArgSerGlyProGluSerGluGlyAlaAlaLeuAlaPro	1118
890	Db	 GlnGlnArgGlySerGlyValArgSerGlyProGluSerGluGlyAlaAlaLeuAlaPro	909
1119	QY	AACCCAGCCCAACTCTGCTCGTCAAGCGCCTCTGGTA---       GlyProProValLeuPheValValAlaValAlaValAlaValProAlaGlu-----	1175
910	Db	 GlyProProValLeuPheValValAlaValAlaValAlaValProAlaGlu-----	926
1176	QY	CCAGGCGCGCATGGCCAA-----       GlyArgAlaGlyGluProLeuValLeuAlaValProGlyAlaAlaGlyProGly	1220
927	Db	 GlyArgAlaGlyGluProLeuValLeuAlaValProGlyAlaAlaGlyProGly	945
1221	QY	CGT----- 	1223
946	Db	ArgAlaAlaLeuLeuAlaProLeuGlyArgTrpValArgAlaGlyGlyGlyAla	965
1224	QY	----- -----GTGCTGGCAGCTGGCGAAGACCGCTCAAGCGCCT	1256
966	Db	GlyValAlaGlyGlyAlaGlyGluAlaGlyLeuGlyAlaGlyAlaGlyLeuGlyAlaGly	985
1257	QY	CGCCACGCGCCACCAATGGAGGAGCAGCGCGCTGCTGGGACAGCAACATGCTCATCCA	1316
986	Db	AlaGlyLeuGlyAlaGlyGlyAlaGlyGlyProGlyAlaGlyGluAlaGlyGlyAla	1005
1317	QY	CTTCTGTGAAGACGCGCCCAA     ArgArgArgArgArgArgTrpAspAspGluAlaGlyLeuLeuGlyProGluArgGly	1346
1006	Db	 ArgArgArgArgArgArgTrpAspAspGluAlaGlyLeuLeuGlyProGluArgGly	1025
1347	QY	GTGGCTGTTCTGCAAGTTCTGTCAGCTGCTGCTCTT-----       GlnAlaGlyArgGlyLeuArgGlyProGlyProArgGlyGlyLeuGlyGluProGlyArg	1394
1026	Db	 GlnAlaGlyArgGlyLeuArgGlyProGlyProArgGlyGlyLeuGlyGluProGlyArg	1045
1395	QY	GCTGTGTTCTGGCGCGCGGTGCGCGGCAAGCAGTACGCGCTGATCAAGCGCAGCGCAT	1454
1046	Db	 GlyHisValGlyArgGlyGluGluGlyArgGlyVal----- GlyProGlyGlyLeuAla	1063
1455	QY	CCCCATTGAGAATCATCGCGCGCACCACTGAGACGCGTGCGCGA       GlyAlaGlyProValHisAlaValAlaHisGlnArgArgHisGlyAlaGlyAspGluGly	1499
1064	Db	 GlyAlaGlyProValHisAlaValAlaHisGlnArgArgHisGlyAlaGlyAspGluGly	1083
1500	QY	----- -----GNACTCGCACGTCGCCACAGCAGAACTACTTCTACTACAA	1538
1084	Db	AspArgValArgGlyLeuProProLeuGlyArgAlaGly----- 	1096
1539	QY	CTGCTCTCACGGCAAGTCTCTGCGCGCACAACTG-----     CTGCTCTCACGGCAAGTCTCTGCGCGCACAACTG-----	1592



QY 1189 GCGAAGCTGTGCTG-----GGTGTGTCAGTCCCTGGCGTGGTG 1227  
Db 383 lYAlaAlaAlaGlnArgThrHisArgArgProProGlyCysProArgSerAlaArgAsp 403  
QY 1228 CTGGAGCTGGGCAAGAC-----CGTAAGGCGCTCCCAACGGCCCAATGAGGAG 1281  
Db 403 roGlyCysProArgThrTrpArgArgArgSerGlyAlaGlnArgGlyHis-----P 420  
QY 1282 CAGGCGCTGTGGGACAGCAACATCTCATCCATTCGTGAAGAACGGGCCCAAGCCG 1341  
Db 420 roProProGlyAlaGlyGlnArgProSerGlyProThrGlyGlyArgProAlaAlaProG 440  
QY 1342 CTGGTGTGGCTGTTCGTCAAGTTCGTAGGCTGTGTCTTCAACAGGCGCTGTCTGG 1401  
Db 440 lYAlaProGlyThrProAlaAlaProGlyProGly----- 451  
QY 1402 TTCGGCGGCGGTGCGGGCAACAGTACGCGTGTATCAAGGGCGGAGCATCCCATTT 1461  
Db 452 -----GlyGlyAlaAlaVal-----ProSerGlyAlaThrProHisP 464  
QY 1462 GAGAACTACATCGCGCGCACCATGACGGCGGTGGCGAGAACTCGCACGTGGCAAGCAG 1521  
Db 464 roGlu-----ArgGlySerGlyProAlaAspProProAla - 475  
QY 1522 AACTACTTCTACTACAACTGCTCACCAGTTCCTGGCGACAACTGCCCCACCTAC 1581  
Db 476 -----AlaAlaArgLeuPro-----ProGluArgGlnGluProArgLeuP 489  
QY 1582 CTGC-----CGAGGGCGGCTTCGCCACCTCAAGAGTGGCGTGGTGA 1625  
Db 489 roGlnAspLeuAlaAlaGlnArgCysProAlaGlyPro----- 502  
QY 1626 CAACCTGACCGCTCCACCACTTCTTCA-----TGGAGGAGCTCAAGCGCG 1673  
Db 503 -----ProProThrArgSerGlyAlaAlaGlnArgThrHisArgArgP 518  
QY 1674 CACCTACACCAAGGTGATCTGTGACCACTGAGTGTGGTATGCGCGTGGGCCAA 1733  
Db 518 roProGly-----CysProArgSerA 525  
QY 1734 CGAGCTGGCGAGTGTGCGGCAAGCAGGTGTGGCGGGCGGATCTGTCTATCTGGCGCTC 1793  
Db 525 laArgAsnProGlyCysProArgThrTrpArgArgArgSerGlyAlaGlnArgGlyHisP 545  
QY 1794 CGCTCCCTCAGCGCGCCCTACGCGAGCTGATCCAGAACGGCGCTTCGACGTGCG--- 1850  
Db 545 roProProGlyAlaGlyGlnArgProSer-GlyProThrGlyGlyArgProAlaAlaPro 564  
QY 1850 ----- 1850  
Db 565 GlyAlaProGlyThrProAlaAlaProGlyProGlyGlyAlaAlaValProSerGly 584  
QY 1851 -----CTGCATCCGCGCGCACTCAGGGCTACATGACGCGGTCAACATGTACAGCTC 1904  
Db 585 AlaThrProHisProGluArgGlySerGlyProAlaAspProProAlaAlaAlaArgLeu 604  
QY 1905 CTCTACATCGCGCGGAGGGCGCCAGAGAGGACAACTA 1946  
Db 605 -----ProProGlu---ArgGlnGluProArgLeu 613

RESULT 15  
T45462

membrane glycoprotein [imported] - equine herpesvirus 1  
C;Species: equine herpesvirus 1  
C;Date: 31-Jan-2000 #sequence\_revision 31-Jan-2000 #text\_change 09-Jul-2004  
C;Accession: T45462  
R;Kirisawa, R.; Kobayashi, T.; Kawakami, Y.; Iwai, H.  
J. Equine Sci. 7, 79-87, 1996  
A;Title: Nucleotide sequences of open reading frames 1, 24 and 71 of an attenuated equine herpesvirus 1  
A;Reference number: Z22973  
A;Accession: T45462  
A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA  
A;Residues: 1-866 <IR>  
A;Cross-references: UNIPROT:O39781; UNIPARC:UPI00000ECBA1; EMBL:D88733; PIDN:BAA20037.1  
A;Experimental source: strain HH1  
C;Genetics:  
A;Note: ORF71  
C;Superfamily: equine herpesvirus glycoprotein X; equine herpesvirus 1 glycoprotein homolog  
F;558-866/Domain: equine herpesvirus 1 glycoprotein homolog <HG>

Alignment Scores:  
Pred. No.: 2,18e-06 Length: 866  
Score: 245.50 Matches: 119  
Percent Similarity: 39.3% Conservative: 76  
Best Local Similarity: 24.0% Mismatches: 204  
Query Match: 6.8% Indels: 97  
DB: 2 Gaps: 13

US-10-620-914-44 (1-1947) x T45462 (1-866)

QY 473 ACAACGTCTACGACACGAGCTTGTCTGCTACCTGCTCCCAAGACGGCTGTGGCGTTGCCG 532  
Db 46 ThrThrAsnSerSerSerProThrThrSerProProThrThrSerSerSerProPro 65  
QY 533 ACTTCTACGTGACGGCAAGTACGACCTGCCCTGCGCCAGATGCCCTGTGCGCGCTT 592  
Db 66 ThrSerThrHisThrSerSerProSerSerThrSerThrGlnSerSerThrAlaAla 85  
QY 593 TCTTCTGCGCATCGATCTTCGACATCGACATCGACAACTTGCATCGCCCGAGCGCGCT 652  
Db 86 ThrSerSerSerAlaProSerThrAlaSerSerThrThrSerThrThrSerThrSer 105  
QY 653 ACCTGAGCAGAGAGCTGAGAGCGGTGTGGGAGGAGAACACCCAGAGGTTCGATCCCTACG 712  
Db 106 ThrGluThrThrThrThrProThrAlaSerThrThr-----ThrProThr 121  
QY 713 TGGCGTGGTGGCGCGCCCTACTACTGCTGTGAGATGCGCGCCCTGCGCCAGGTGGCC--- 769  
Db 122 ThrThrThrAlaAlaProThrThrAlaAlaThrThrAlaValThrThrAlaAlaSer 141  
QY 770 -----AGCCCTGCGACGAGGCGGTGAGCGCGCGCGCG 802  
Db 142 ThrAlaAlaSerThrSerAlaGluThrThrThrAlaThrAlaThrAlaThrSerPro 161  
QY 803 CCATGTTTCCGCGCCCTCTCTGTACACGACGCTGTGGAGGAGACCCGAGCCGATATGG 862  
Db 162 ThrThrThrThrProThrSerThrThrThrThrAlaThr----- 176  
QY 863 AGGTGATGAGATCAACCCCAAGGACACGGTGTGACCCCTGTAGCTAGCGCGCGCTGCAATG 922  
Db 177 -----ThrValProThrThr----- 181  
QY 923 CCCTGAACCTGTGTTGTCAGGCGCGCGAGTGTGTGCTGGTGTGACTGCAACCCCGCGC 982  
Db 181 ----- 181  
QY 983 AGTGGCGCTTCTGGAGCTGAAGAAGGTGGCCATTCAGCAGCTGGAGTTTGAGACGTGT 1042  
Db 182 -----AlaSerThrThrThrAspThrThr 189  
QY 1043 GGCAGCTGTTCGCGAGGCGGTGCACCCGCGCATGAGGAGCTGTACGAGAGAGCTGG 1102  
Db 190 ThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrAlaAlaThrThr 209  
QY 1103 CGCCCTTCTGTGCGCAACACCGACCACTTCTGGTCCAGCGCCCTCTGTACTTCCAGC 1162  
Db 210 ThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrAlaAlaThrThr 229  
QY 1163 ACGGCTGTACTACAGGCGCGCATGGCAAGCTGTGCTGGGTGCTGCTGCTGCTGCGCG 1222  
Db 230 ThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrSerSerAlaThr 249  
QY 1223 TGGTGTGGGACTGGGCAAGACCGTCAAGCGCTCGCCCAACGCGCCCAACCAATGGAGGC 1282

```
Db      250 AlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThr----- 263
Qy      1283 AGCGCCGCTGTGGGACAGCAACATGCTCATCCACITTCGTGAAGAAGCGGCCCAAGCCGC 1342
Db      264 -----ThrAlaAlaThrThrThrAlaAlaThrThr-----ThrAlaAlaThrThr 278
Qy      1343 TGGTGTGGCTGTTCGTCAGCTGCTGAGCCCTGCTGCTCTTCAACAAGGCGGTGTGTGGT 1402
Db      279 ThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrAlaAlaThrThr 298
Qy      1403 TCGGCGCGCGCGTCCGGGCAAGCAGTACCGGCTGTATCAAGGCGGAGCGGATCCCCATTG 1462
Db      299 ThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAla----- 316
Qy      1463 AGAATACATCGCGCGCACCATGAGCGGCTGGGGAGAACTCGCAGCTGGCGAAGCAGA 1522
Db      317 ---ThrThrThrAlaAlaThrThrThrAlaAlaThrThrThrGlySerProThrSerGly 335
Qy      1523 ACTACTTCTACTACA-----ACTGCCTCAACGGCAAGTTCTCGCGGCAACTGCC 1573
Db      336 SerThrSerThrThrGlyAlaSerThrSerThrProSerAlaSerThrAlaThrSerAla 355
Qy      1574 CCACCTACCTCGCGAGCGGCTTCGCCACCTCAAGAGTGGGCTGGTGGAACCTGA 1633
Db      356 ThrProThr-----SerThrSerThrAlaAlaAlaThrThrSerThr 370
Qy      1634 CGGTCTCCCAACT-----TCTTCATGGAGAGCTCAAGCGCGCACCTACACCAAGG 1687
Db      371 ProThrProThrSerAlaAlaThrSerAlaGluSerThrThrGluAlaProThrSer--- 389
Qy      1688 TGATTCTGATGGACCACTGGAGCTGGCTGGATATGCCGTGGCCACAGAGTGGCGCGAGT 1747
Db      390 -----ThrProThrThrAspThrThrThrProSerGluAlaThrThrAlaThr 405
Qy      1748 GCCTGGCCAAAGCAGGTTGCGCGGCGGCGCATCGTCAATCGCGGCTCGCCTCCCTCAGCC 1807
Db      406 ThrSerProGluSerThrThrValSerAlaSerThrThrSerAla---ThrThrThrAla 424
Qy      1808 CGCCTACGCGGAGCTGATCCAGAGGCGGCGGCTTCGACGTGCGCTGCATCGCGCGGCCA 1867
Db      425 PheThrThrGluSerHisThrSerProAspSerSerThrGlySerThrSerThrAlaGlu 444
Qy      1868 CTCAGGCTACATGAGCCGCTCAACATGTACAGCT---CCTTCTACA 1912
Db      445 ProSerSerThrPheThrLeuThrProSerThrAlaThrProSerThr 460
```

Search completed: March 14, 2006, 01:45:57  
Job time : 82 secs

Result No.	Query			ID	Description
	Score	Match	Length		
C 1	374.5	10.3	663	US-09-253-991A-30843	Sequence 30843, A
	352	9.7	310	US-09-248-796A-21233	Sequence 21233, A
	332.5	9.2	563	US-09-252-91A-30843	Sequence 30843, A
C 4	332	9.2	536	US-09-252-991A-16754	Sequence 16754, A
C 5	326.5	9.0	681	US-09-252-991A-24567	Sequence 24567, A
C 6	324	9.0	2294	US-09-253-991A-17231	Sequence 17231, A
C 7	323	8.9	638	US-09-253-991A-27068	Sequence 27068, A
C 8	317.5	8.8	467	US-09-252-991A-18296	Sequence 18296, A
C 9	316.5	8.7	1706	US-09-252-991A-31760	Sequence 31760, A
C 10	310	8.6	1073	US-09-252-991A-27341	Sequence 27341, A
C 11	309	8.5	798	US-09-253-991A-23774	Sequence 23774, A
C 12	308	8.5	1064	US-09-252-991A-17508	Sequence 17508, A

Db 4 ProAlaProProAlaTTrp-----ProGlnArgTrpProProArgAlaAla 19

1717 CCAGCCAGTCCA-----CGTGGTCCATCAGATCA----- 1688  
Db  
20 ArgAlaSerProTrpSerThrThrArgProProAlaAspArgSerGlyAlaThrAla 39  
1687 -----CCTTGGTGT-----AGTGGCGGCTTTGAGTCTCT-----CCA 1655  
Db  
40 ArgValProAlaCybArgArgValProIleArgCysAlaAsnAlaTrpProGlyArgPro 59  
1654 TGAAGAAGTTGGTGAGACCGGTGTCACACCGCCACTCTTGAGG----- 1604  
Db  
60 MetSerSerThrSerArgArgProGlyTrpSerProAlaValProAlaGlyAlaCysCys 79  
1603 TGGCGAAGCGCGCTCCGCGAGTAGG-----TGGGCGAGTTGTCGGCGCA 1559  
Db  
80 TrpLysThrArgAsnValAlaGlyArgSerAlaThrAspAlaTrpSerSerAlaProAla 99  
1558 ---GGAACCTTTCGGGTGAGGCGAGTTGT-----AGTAGA 1529  
Db  
100 ProAlaAsnCys-----CysCysLeuPheProAlaGlyProSerProAlaSerPro 116  
1528 AGTAGTTCTGCTTCGGCAGCT-----GCGAGTTCTCGGCCACCGCTCCATGG 1481  
Db  
117 AlaProAlaAlaCysArgArgTrpProArgAlaAlaCysHisTrpProAlaSerAlaTrp 136  
1480 TGC-----GCGCGATGTAGTTCTCAATGGGGATGCCGTCCG---CCTTGATCA 1436  
Db  
137 TrpTrpLeuAlaProAlaArgCysCys-----TrpProAlaProAlaProAlaSer 154  
1435 GCGGTACTGCTTGC-----CGGSCACGCGCGCGCGAACCACGACGCGCTTGT 1385  
Db  
155 AlaGlyArgAlaCysCysAlaSerProAsnArgArgArgGluProTrpProPro--- 173  
1384 TGAAGACACAGGCTCACGAACCTTGACGAACGCCACACGCGGCTTGGCGCGTCT 1325  
Db  
174 -----SerProTrpAla----- 177  
1324 TCAGAGTGGATGAGCATGTGTGTGCCACAGCGGCTGTCTCTCCATTGTGGCG 1265  
Db  
178 SerArgAlaGlyProAlaSerCys-----GlyArgProPro----- 189  
1264 CGTTGGCGAGCGCTTGACGGTCTTGC-----CCAGTCCCACGACCGCGCCAGCTGCA 1208  
Db  
190 -----AlaCysSerProValAlaThrAlaProThrAlaThrCys 202  
1207 GCACCCAGCAGCT-----TGCCCATGCCGCGCTGTGTAGTACAGGC 1166  
Db  
203 SerProProSerAlaArgSerAlaTrpLysProCysAlaCysAlaLys---AlaValGly 221  
1165 CGTGTGGAAGTACCAGAGCGCTTGACCAAGAGTTGTGGCTGTTGGCAGAGGAAG 1106  
Db  
222 SerAlaArgSerProAlaAsnAlaTrp----- 230  
1105 GCGCCAGCTTCTTCGTACAGCTCTCAATGCGCGGGTGACGCGCTCGCCGCAACAGCT 1046  
Db  
231 ---ProAlaAlaSerAlaTrpCysProThrCysAsnSerAlaArgProSerAlaIleAla 249  
1045 GCACCA-----CGTCTCT-----CAACT 1028  
Db  
250 SerThrAlaArgArgSerProSerThrAsnGlyArgProAlaAlaCysProThrThr 269  
1027 CCAGCTGCTGAATGGCCACCTTCTTCAGCTCCAGAACGCGGCTGCGCGGGTTGCGAGT 968  
Db  
270 ProProAlaAsnAlaProAlaSerAlaAlaAsnTrpProTrpTrpLysAlaProSer 289  
967 CCACCGACACACCTGGCGCGCCCTCGACCA---GCAGTTTCAGGCGCATTCGACCGCG 911  
Db  
290 ProAlaMetProProSerThrSerAlaThrArgLeuAlaAlaCysGlyHis----- 306  
910 CGTAGTCAGGTCAGCA-----CCGTGT 887  
Db  
307 ---AspGlyAspAlaGlyArgAlaSerGlnThrProTrpArgAspThrSerProCys 324  
886 -----CCTTGGGGT-----TGATCT 872

325 ValProAsnCysValAsnTrpArgArgArgIleProTrpSerAlaValAlaLysThrSer 344  
QY  
871 CMATCACTTCATATCCGCT-----CGGGTCTCT---CCCACGACTCGGTACAGGA 821  
Db  
345 ProTrpProArgTrpProGlyMetProAlaGlyProArgProSerCysThrAlaValAla 364  
QY  
820 AGTGTGGCG-----GGAACATGGCGCGCTCCACGC---GCTCTCTGTCGAGGG 773  
Db  
365 AlaTrpAlaProAlaArgGlyGlyTyrAlaAlaProArgAsnSerCysSerAlaGly 384  
QY  
772 CGTGCCCAACGC-----TGGGCGAGCGGC 749  
Db  
385 ArgLeuArgArgValArgArgSerAlaArgAlaTrpArgProTrpProAlaGly 404  
QY  
748 CAATCACA-----CGTAGTAGGGGCGCGCGCCACG 716  
Db  
405 ArgAlaThrProAlaSerAspAlaArgAspSerAlaSerArgGlyArgProAlaThr 424  
QY  
715 GCAGGTAGGGATCGAACCTGGGTGTCTGTCTCCACACGCGCTCCAGCTTCTGTCTCCA 656  
Db  
425 AlaAlaAspHisProAlaAlaTrpValSerAla---AlaArgArgThrSerSerAlaPro 443  
QY  
655 GSTAGGCGCGCGCTCGGGCCGATGTCAATGTGTGATGTGCAAGATCGATCGCCAGA 596  
Db  
444 IleAlaGlySerAlaProGlySerGlyThrAlaProArgCysHis----- 458  
QY  
595 AGAACCGGCGGACCGAGGCATCTGGCGAGGGCAGGTCTGTTCCGCTCAGCTCAGGTAGA 536  
Db  
459 -----ProValArgLysAspGlyAlaGlyProAlaSerThrGlyArgSerArgArg 475  
QY  
535 AGTCGCAACGCCACCGCGCTCTGGGACAGGTACGAGCAAGCTGCTCGATGACGT 476  
Db  
476 ArgTrpAlaAlaAlaProAlaArg-----ArgAlaArg 486  
QY  
475 TGTGAACCGTGGAAATCATCTGAGCGAGTAGGAGAAGTGTAGAGGTGCGGTGCGCT 416  
Db  
487 ThrGlyLe-----ArgArgThrAlaAlaAlaCysTyr 498  
QY  
415 CAGGGGCGCMAATGGCAAGCTCGGCTCCA----- 383  
Db  
499 ProProProAlaProGlyArgArgThrProAlaAlaArgArgSerAlaArgThrAla 518  
QY  
382 -----CGACCTGGACATTTCTCCAGCCCTTCCGCTTCCGCTTCCGCT 347  
Db  
519 ProArgProAlaIleProAlaHisArgProSerAlaSerAlaThrGlyTrpProGlyPro 538  
QY  
346 TCTTCTTGGCCACTCGCACGAGGTGGCACAGGTGCGACCGTAGATGGAACCT 287  
Db  
539 AlaLeuAlaCysProAlaAlaAla----- 546  
QY  
286 TCGCCAGGTGATGTAATCAGCCATCATATCGACATTTCTCCAGTCCAC---CACCCA 230  
Db  
547 -----GlyArgArgProSerProAlaProAlaProAlaProAlaProAlaArgPro 564  
QY  
229 GGTCAACCCAGATGAGTTTCGAGCGCTCGGCCGCGGCGAGCAAGCGCGCTGGG 170  
Db  
565 ArgAlaProArg-----ArgArgProGlyArgSerProAlaThrAlaProAla 580  
QY  
169 GCGCGTAGAAGCTTCTCAGGCGAGCAGCGTGTATCATCGCTTCTTCTGTCGCCAACACA 110  
Db  
581 Ala---PheAlaAlaProAlaGlyProThrProAlaArgProProAlaAlaAlaThr--- 598  
QY  
109 TATGGCCGACAGCGGTGAGTCTCTTTCATGTGTGCTGCTGAGCTTCTCCAGGGAGA 50  
Db  
599 -----GlyArgProSer-----AlaProProArgProAlaArg 609  
QY  
49 AGTTCT-----TCTGTGTAGTCTCGCAGCGCGCTCAGCAGCCGCCACCCA 2  
Db  
610 ArgArgValArgTrpProTrpArgSerAlaArgCysSerProArgArgProThrPro 628

RESULT 2

US-09-248-796A-21233



```

868 ATGAGAGATCAACCCCAAGGACACGGTGCTGTGACCTCTGACTAGCGGCGGTGCAATGCCCTTG 927
   :: ||||||||| ::||| ||| ||| ||| |||
215 LeuAsnPheThrSerAspAspThrValLeuAlaIleThrSerAlaGlyAspAsnIleLeu 234
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
928 AACCTG-----CTGGTGCCAGGGCGCGGCGGCTGTGTGGTGTGACTGCACACCCCGCG 981
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
235 SerTyrAlaSerLeuProThrProProlYsIleHisAlaValAspLeuAsnProCys 254
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
982 CAGTCGGCGCTTCTGGAGCTGAAGAAGGTGGCCATTACAGCAGCTGGAGTTTGAGGACGTG 1041
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
255 GlnAsnHisLeuLeuGluLeuLysLeuAlaSerPheArgCysLeuSerGlnGluGlnIle 274
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1042 TGGCAGCTGTTCCGGCAGGGCGTGCACCCCGCGCATTTGAGGAGCTGTACGAGAAGAAGCTG 1101
   ||| ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
275 TrpSerMetPheGlyGluGlyIleGluAsnPheAsnAspLeuLeuIleAspThrLeu 294
   ||| ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1102 GCGCGCTTCTGTCGGAACACCCAGCACAACTTCTGGTCCAAGCGCTCTGTACTTCCAG 1161
   ||| ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
295 AlaProHisMetSerSerAsnAla-----PheGln 304
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1162 CAC 1164
   |||
Db 305 His 305

RESULT 3
US-09-252-991A-30843
; Sequence 30843, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30843
; LENGTH: 663
; TYPE: PRN
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30843

Alignment Scores:
Pred. No.: 2,81e-16 Length: 663
Score: 332.50 Matches: 193
Percent Similarity: 32.4% Conservative: 52
Best Local Similarity: 25.5% Mismatches: 253
Query Match: 9.2% Indels: 259
DB: 2 Gaps: 44

US-10-620-914-44 (1-1947) x US-09-252-991A-30843 (1-663)

Qy 29 CGAGCTACACAGA-----AGAACTTCTCCTGGAGAGCTCA----- 67
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12 ArgTrrProProArgAlaAlaArgAlaSerProTrrSerThrThrArgProPro 31
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 68 -----AGTCCACAGCATGAAGGATGACCTGACCGTCTTCGGCCATATGTGTTGCG 118
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 32 AlaAspArgSerGlyAlaThrAlaArgVal-----ProAlaCys----- 44
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 119 GCACGAAGAAGGCGGATGATCAGCTGTCTC-----GCTGTG-----AGAGCTTCT 163
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 45 -----ArgArgValProIleArgCysAlaAsnAlaTrrProGlyArgProMetSer 62
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 164 ACGGGCCCCAGCGCGCTTGTGTCGGCGCTGCCGAGCGCTCGAAGCTCATCTCGG 223
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 63 ThrSerArgArg-----ProGlyTrrSerProAlaValProAlaGlyAla 77
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 16754  
; LENGTH: 536  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-16754

## Alignment Scores:

Pred. No.: 2,8e-16 Length: 536  
Score: 332.00 Matches: 181  
Percent Similarity: 33.1% Conservative: 34  
Best Local Similarity: 27.9% Mismatches: 193  
Query Match: 9.2% Indels: 241  
DB: 2 Gaps: 36

US-10-620-914-44 (1-1947) x US-09-252-991A-16754 (1-536)

QY 1888 CGGGTCCATGTAGCCCTGAG-----TGGCGGGCGGATGCGAGCGCACTCGA 1841  
DB 8 ArgArgProSerProSerSerAlaSerAlaTrpSerAlaAlaSerSerMetPro 27  
QY 1840 AGCCCG-----CCTTCTGGATCA 1823  
DB 28 SerProAlaAlaAlaArgLysProAlaSerThrSerSerAlaSerAlaSerAlaSer 47  
QY 1822 GCTCGCGGTAGGCGGGCTGAGGGAGCGG-----AGCCAGATGACGA----- 1778  
DB 48 AlaAlaArgArgPro-----GlyArgArgProIleAlaCysAlaAlaArgTrpArgArgSer 65  
QY 1777 TGC CGCGCGCGCAACCTGCTTGGCCAGGCACTCGGCCAGCTGTTGGCCAGGGCATAT 1718  
DB 66 CysArgProThrProSerCysTrpLysProMetArgArgThrTrpArgHisArgCysIle 85  
QY 1717 CCAGCCAGTCCA-----CGTGGTCCATCAGAAATCACCCTGGTGT 1679  
DB 86 ProAlaSerAlaThrValArgSerThrCysArgThrSerAlaAlaProSerProSerCys 105  
QY 1678 AGTGCGCGCTTGAGCTCTCCATGAGAGAGTTGGTGAGAGCGGTACAGTTGTTCCACCA 1619  
DB 106 -----AlaGlySerThrPro 110  
QY 1618 CGCCACTCTTGAGGGTGGCGAAG-----CCGCTCGCGCAGGTAGTGG 1574  
DB 111 ArgAsn-----TrpArgArgAlaAlaAlaThrProPro----- 122  
QY 1573 GGCAGTTGTGCGCGAGAACT-----TGCGGTGAGCGAGTTGT----- 1535  
DB 123 ---SerCysSerAlaGlyThrAspGlnAlaSerArgThrSerGlyAlaSerCysGly 141  
QY 1534 -----AGTAGAAGTAGTTCTGCTTGGCGAGTGGCGAGTTCTCGGCCA 1493  
DB 142 TyrValSerArgAsnAlaSerArgGlnAlaAspSerCysAla---AlaSerSerGlyArg 160  
QY 1492 CGCGGTCCATGTTGCGCGGATGTAGTTCTCAATGGGGATGCGTCCGCTTGATCAGCG 1433  
DB 161 ArgArg-----Ala 163  
QY 1432 CGTACTGTGTCGCGCAGCGCCGCCGCAACCAACGACGCGGCTTGTGTAAGAGCACCA 1373  
DB 164 LysSerLeuGlnProCysArgArgAsnArg-----ArgGlyIle 176  
QY 1372 GGTCTACGAACCTTGACAGACGACCAACCGCGGCTTGGCGGCTTCTTCACGAAGTGA 1313  
DB 177 GlyThrArgProLeuMetAlaSerGlnLysAlaSerGlnLysProThrHisGlyGly 196  
QY 1312 TGAGCATGTTGTTGCCACAGCGCGCTGC-----TCCT 1277  
DB 197 --AsnAlaArgAlaSerArgArgSerCysLysArgAlaAlaLeuSerThrSerAlas 216

QY 1276 CCATTGTGGCGCGTGTGGCGAGCGCTTGACGGTCTTGC-----CCAGTCCAGCA 1226  
DB 216 e-SerMetAlaAlaArgLeuAlaAlaArgPheSer-SerSerMetGluProProAla 235  
QY 1225 CCACGCGCAGGCACTGCGAGCACCAGCACAGCTTGGCCCTGCTAGTACAGGC 1166  
DB 236 MetArgArgAsnAlaAlaArgThrArgThrArgAlaArgGlySerProGlyGlySerThrGly 255  
QY 1165 CGTGTCTGGAAGTACCAGAGCGCGCTTGGACAGAAAGTTGTGGCTGTTTGCACAGGAAG 1106  
DB 256 ArgGlyGlyThrArgGly----- 262  
QY 1105 GCGCCAGCTTCTCTGTAAGCT---CCTCAATGCGCGGTGCA----- 1064  
DB 263 SerSerAlaSerProArgProAlaThrProAlaAlaAlaGlyAlaAlaArgAla 282  
QY 1063 -----CGCCCTCGCGACAGCTGCCACAGCTCTCAAACTCCAGCTGCTGAA 1016  
DB 283 ArgProArgProArgProAlaArgThrCysArgProAlaProCysArgProGly----- 300  
QY 1015 TGGCCACCTTCTTCACTCCA-----GAAGCGCGACT 983  
DB 301 -----ProSerAlaAlaProArgArgHisArgArgArgAspArgArgProSer 318  
QY 982 GCGCGGGTGTGAGTCCACCG-----ACACCACTGCGCGGCCCTGCA----- 938  
DB 319 GlySerGlyCysGlyProAlaGlyThrProSerGlyArgValProGlyArgArgVal 338  
QY 937 -----CCAGCAGCTTCAGGGCATTCAGCGCG 911  
DB 339 GlyArgAlaGlyAsnCysAlaAlaGlyArgProAlaAlaGlyAlaAlaArgAlaArg 358  
QY 910 -----CGTAGTCAGGTCAGCACCGCTCTTGGGTTGATCTCCATCA 866  
DB 359 ProGlyArgProValArgArgArgAlaLeuAlaGlyCys----- 371  
QY 865 CTCCATATCCGGCTCGGGTCTCTCCACGACTCGCTGTACAGAAAGTGGGGGGAACA 806  
DB 372 ---ProValProAlaAlaGlyArgProGlyTrpSerArgProAlaAlaArgSerGly--- 389  
QY 805 TGGCGCGCGCTCCAGCGCTCTCGTGCAGGCGCT-----GGCCAACTGCTGG 758  
DB 390 -----SerProArgProValAlaGlyArgAspProSerProGlyArgArgAla 406  
QY 757 GCAGCGCGCAATCCACACGCTAGTAGGGGCGCGACCGCACGCGCATAGGGATCGAAC 698  
DB 407 AlaGlyAlaGlySerValAlaAsp----- 414  
QY 697 CTGGGTGTCTGCTCCACACGCTCCAGCTTCTGCTCCAGGTAGGCGCGCGCTCGG 638  
DB 415 -----SerAlaProGlyThrArgArgAlaArg 423  
QY 637 GCGCGATGCAATGTTGTCGATGTCGAAGA-----TCGATCGCCAGAGAAGAAC 590  
DB 424 AlaAlaAlaGlyCysGlyArgSerArgArgAlaAlaAlaValProArgArgSer 443  
QY 589 GCGCGCACCGAGGCTCTGGCGCAGGGCAGTCTGACT---TGCCGCTCACGTAGAGT 533  
DB 444 GlyProCysArgProAlaAlaThrGlyAlaProProProGlyCysAlaThrArgArg--- 462  
QY 532 CGGCAACGCCACCGCGCTTGGACAGGTACGAGCAAGCTCGTGCATGACGTTGT 473  
DB 463 -----GlyHisAlaGly---GlyThrArgArgProArgGluSerGlyAla 476  
QY 472 GGAACGGTGGATCATCTGAGCGAGTAGAGAGGTGATGAGCTCGCGGTGCTCCCTCAG 413  
DB 477 Gly-----CysGlyTrp 480  
QY 412 GGGCGCGCAATTTGGCAAGCGTCCGCTCCACGA-----CCTGGACATTTCTTCAGCCCT 359  
DB 481 SerGlyArgThrGlyVal-----ProProAlaArgPheProGlyAlaSerAlaProArg 498

```

QY 358 TGGCCCTTGGCTTCTTCTTGGCCACCTCGCACAGCGAGTGGCACAGGTGCACACCGTAGA 299
Db ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 499 TTPProProProGlylySerProThrGlyAlaArgArgThrThrSerArgCysArg 518
Db ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 298 TGGACTTGAACCTTCGCCAGGT 278
Db ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 519 TTPCysGlyArgThrProGly 525
Db ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 5
US-09-252-991A-24567
; Sequence 24567, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24567
; LENGTH: 681
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24567

Alignment Scores:
Pred. No.: 7,88e-16 Length: 681
Score: 326.50 Matches: 214
Percent Similarity: 32.8% Conservative: 38
Best Local Similarity: 27.8% Mismatches: 262
Query Match: 9.0% Indels: 255
DB: 2 Gaps: 50

US-10-620-914-44 (1-1947) x US-09-252-991A-24567 (1-681)
QY 1873 CCTGAGTGGCGCGCGGATGC---AGCGCAGTTCGAAGCCCGCT-----TCGTGA 1826
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1825 TCAGCTCGGCTAGGGCGCGCTGAGG-----AGCGGAGCGCCAGA 1784
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1783 TGACGATGCCCGCGCGCAACCTGCTTGGCCAGC-----ACTCGGCCA----- 1739
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 48 AsnArgSerAsnArgAlaTTPProArgTTPProGlyCysAlaArgThrAlaProThrSer 67
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1738 -----GCTCGTTGG---CCACGGGATATCCAGCCAGTCCA 1706
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 68 AspCysTrpSerAlaAsnAlaAlaSerArgTrpSerIysArgSerAlaProAlaPro 87
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1705 CGTGCTCATCAGNATCACCTTGTGTAGG-----TCCGCGCTTGGCTCTCTCCA 1655
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 88 ArgAlaSerTrpSerThrArgThrAlaAlaGlySerProCysAla----- 102
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1654 TGAAGAAGTTGGTGACAGCTCAGGTTGTCCACCGCCACTCTTGAGGCTGGCGAAGG 1595
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 103 -----TrpMetProIysAlaAlaCysArg-----ArgArg 112
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1594 CCGCTCCCGCAGGTAGTGGGCGAGTTGTTCGCGCAGGAACCTTCGCGTGAAGCAGTTGT 1535
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 113 ProAlaAlaAlaMetArg-----SerCysArgTrpSerSerGly 125
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1534 AGTAGAAGTAGTTCTGCTGGCAGCTCCGAGTTCTCGG---CCACCGCTCCATGTGTC 1478
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 126 AsnGlyArgTrpArgTrpArgArgAlaValArgProTrpProAlaSerProArgVal 145
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1477 GCGGATGTAGTTCTCAATGGGATGCGCTCGGCTTGTATCAGCGCGTACTGTCTTGCCCG 1418
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db ||||| :||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 146 AlaArgValAlaGlyValTrp-----ProPro-----SerLeuCysPro 158
QY 1417 GCACGCGCGCGCGGACCAACACACAGCAGCGCTTGTGAAGAGCACCAGCTCAGCACTTGA 1358
Db ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 159 AlaProProArgArg-----ArgArgProGly---ArgGlnArg 170
Db ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1357 CGAACAGCCACACACAGCGCTTGGGCGCTTCTTTCACGAAGTGATGAGCATGTTGCTGT 1298
Db ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 171 ArgProArgGlyProAlaAlaGlnArg----- 180
Db ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1297 CCCACAGACGCGCTCTCTCCATGTGGCGCGTTCGCGAGCGCCTTCAGCTCTTGC 1238
Db ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 181 -----AlaAlaTrpArgArg-----ArgAlaGly 188
Db ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1237 CCAGTCCAGCAGCAGCGCGC-----ACT 1211
Db ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 189 AspGlnProAlaAlaArgProGlyArgGlyGlnArgThrCysLeuGlnProLeuLeuAla 208
Db ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1210 GCAGCAGCAGCAGCACA---GCTTGCCCATGCCG----- 1181
Db ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 209 ValGlnProAlaValLeuGlnCysProAlaArgAlaGlyAspHisSerArgArgGly 228
Db ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1180 -----CCTGGTAGTACAGGCCGT 1163
Db ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 229 SerGlyGlyThrGlyAspProSerArgProGlyArgGlyAspGlyProProGlyGlu 248
Db ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1162 GCTGGAAGTACACAGCGCTTGGACAGAGTTGTGGCTGTTTCCGACAGCAGGCGC 1103
Db ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 249 ProGlyAlaAspArg-----LeuAspSerGlyGly 259
Db ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1102 CCAGCTTCTTCTGTACAGCTCT-----CAATGCGCG 1070
Db ||||| :||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 260 ProAlaHisAlaProAlaAlaProThrProArgLeuHisArgAlaArgGlnSerLeu 279
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1069 GGTGCAAGCTCGCGCAACAGCTGCCACAGCTCTCAAACTCCAGCTGTGTAATGGCCA 1010
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 280 AlaProArgProArgArgIleProProArgSerArgGlyProAlaAlaProLeuPro 299
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1009 -----CCTTCTTCAGCTCCAGACCGCGCGCTGCGGGTGC 971
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 300 LeuArgAspProAlaGlyAlaProGlyArgArgProArgLeuThrLeuAlaGlyAla 319
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 970 AGTCCACCGACACACCTGCGCGCGCTCCACAGCAGGTTCAGGCGCATTC----- 917
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 320 Ala-----AlaProProArgArgAlaGlyGlyGlyArgValLeuArg 333
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 916 ---AGCGCGCTAGTTCAGGTCAGCACCGTGT-----CCTTGGGGTTCATCTCCATCA 866
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 334 ArgSerArgArgGlySerGly---LeuProCysLeuArgPro-----ValAlaAsp 349
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 865 CTTCATATCCGCTCGGGTCTCTCCACGACTGCG-----TGTACA 824
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 350 ProThrLeuProAlaThrCysProAlaProGlyAlaArgGlyArgAspGlyHisArgSer 369
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 823 GGAAGTGGCGCGGA-----ACATGGCGCGCGCTCCA 791
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 370 GlyArgArgProGlyGlyArgArgArgGlnProGlyLeuValAlaAlaGly 389
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 790 CGGCTCTCTGTCAGGCGCTGCCAACCGTGGCGAGCGCG----- 749
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 390 ArgAlaAlaArgSerGluCysArgArgAlaAlaGlyHisProGlnProValArgThr 409
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 748 -----CAATCCACACT----- 737
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 410 GlyLeuGlyArgPheArgValGlnSerArgArgProAlaProArgLeuProGlyIle 429
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 736 -----AGTAGGGGCGCGCAGCCACCGCGCAGTAGGGATCGAACCTTGGGTGTCT 686
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 430 ProArgAspAlaAlaGlyGlnProGlyLeuAlaArgAlaThrHisArgProCysAsp 449
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 685 GCTCCACACAGCGCTCCAGCTTCTGCTCAGGTAGGCGCGCGCT---CGGGCGCGATGT 629
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Db 450 GlyProAlaThrSerLeuAlaAsp---ProAlaArgProAlaAlaProAlaArgArgLeu 468  
 Qy 628 CAATGTTGTCGATGTCGAAGATCATGCCAGAGAAACCGCGACCGATCTGGC 569  
 Db 469 ProAlaLeuProAlaThrArgAlaAlaProAla---GlyAlaGlyGlyGlnSerArg 487  
 Qy 568 GCAGGGCAGGTCGATCTGCGCTCAGCGTAGAAGTCGGCAACGCCACCGAGCGCTTT 509  
 Db 488 LeuGlyAlaGlyHisArg---ArgGlyProArgHisArgAlaArgArgProAlaArg--- 505  
 Qy 508 GGGACAGTACGACGACCTGCTCGATGATGCTGTGGAACGGTGGATCATCTGAGCG 449  
 Db 506 GlyThrGlyThrProPro-----SerAlaGlyTyrProArgAlaAlaLeuArg 521  
 Qy 448 AGTAGGAGAAAGGTGATGAGCGTCCGCTGCGCTCAGGGGGCGCAAAATGGCAAGCGTCGG 389  
 Db 522 AlaAlaArgArg-----ThrLeuArgAlaPro---GlyAlaMetAlaGlyArgHis 538  
 Qy 388 CTCCACGACCTGACATCTTCAGCCCTTGGCCTTGGCTTCTTCTTGGCCACCTCGC 329  
 Db 539 -----GlyHisHisGlnHisProArgProAlaGlu-----ProValArg 551  
 Qy 328 ACAGCGAGTGCACAGCT-----CGACCAGTAGATGAGCTTGAATCTGCCCA 281  
 Db 552 LeuValAlaGlyProGlyHisProLeuAlaArgProArg-----ThrPro 567  
 Qy 280 GGTGATGTAATCAGCCATCATATCGACATCTCCCGAGTGCACACCCAGGTCAACCC 221  
 Db 568 GlnArg-----ArgGlyMetArgGlyArgProGlyThrAlaArgArgGlyAlaPro 584  
 Qy 220 AGATGAGTTGAGCGCTCGCGAGCGCGGAGCAAGGCGAGCGCTGGGGCCCGPAGA 161  
 Db 585 GlyProGlyArgGlnProArgThrAlaGlyArgSerArgProArg-----ArgArg 601  
 Qy 160 AGCTCTCCAGCGAGCGAGCTGATCATCGC-----CCTTCTGCTGCCGA 116  
 Db 602 SerThrAlaGly-----ArgLeuHisArgLeuCysArgGlyAsnProArgAlaValGly 619  
 Qy 115 ACCACATATGCGCAGAACCGTCAGGTCTCATCTTCATGCTGC-----74  
 Db 620 AlaThrAlaProGlyArgProGlyGlnProGlyAlaAlaGluProAlaArgProArg 639  
 Qy 73 -----TGAGTTGAGCTTCTCCAGGGAGAAAGTTCTTCTTGGTGTAGCTCG 29  
 Db 640 ArgArgProGluLeuAlaAlaAlaLeuAlaGlyLysArgArgAlaAspAlaArgAsn 659  
 Qy 28 CAGCGCGCGCTCAGACCCGCCCA 2  
 Db 660 AlaAlaGlyArgProAlaProAlaPro 668

## RESULT 6

US-09-252-991A-17231  
 ; Sequence 17231, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252.991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 17231  
 ; LENGTH: 2294  
 ; TYPE: PRT  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-17231  
 Alignment Scores:

Pred. No.: 1,97e-15 Length: 2294  
 Score: 324.00 Matches: 274  
 Percent Similarity: 29.3% Conservative: 43  
 Best Local Similarity: 25.3% Mismatches: 280  
 Query Match: 9.0% Indels: 485  
 DB: 2 Gaps: 58  
 US-10-620-914-44 (1-1947) x US-09-252-991A-17231 (1-2294)  
 Qy 20 GCGGGCTCGAGCTACACCAAGAACTTCTCCCTGG-----58  
 Db 55 AlaGlyHisArgGlnArgProArgArgAlaAlaGlyTyrAlaValGlnHisProArg 74  
 Qy 59 ---AGAAGCTCAAGCTCAGCAGCA-----TGAAGGATG 88  
 Db 75 ArgArgSerGlyAlaGlyAlaGluCyeArgGlyLysProAlaArgTyrHisArgLeu 94  
 Qy 89 ACCTGACGTTTTCGCCCATATCTGTTTCGGCAGCAAGAGGGCGATGATCAGCTGCTC 148  
 Db 95 ProArgPro---AlaAlaGluAlaGlyArgHisProArgHisGly-----108  
 Qy 149 GCCTGGAGAGCTTCTACGGGCCCGCAGCGCTCTGCTGCTCCGCTCGC-----200  
 Db 109 -----ArgGlyArgGlyProArgArgProLeuArg-GlnProGlyAlaGlySe 126  
 Qy 201 -----CGAGCGCTCGAACTCATCTGGGTTCACCTGGTGGTGGTGGCTG 244  
 Db 126 rAlaArgArgGlnCyeArgLeuArgGluProArgProGlyGlyProGly-----142  
 Qy 245 GGAGAAATGTCGATATGATGGCTGATTACGACTGCGCAAGTTCAGTCCATCTACG 304  
 Db 143 -----GlnProGlyValProProAlaAspArgAl 152  
 Qy 305 TGGTCCACCT-----GTCCACTCGCTGCGAGG 334  
 Db 152 aArgArgProArgArgIleArgArgGlyArgAlaGlyProArgGluAlaArgG 172  
 Qy 335 T---GCCAAGAAAGAGCGAAGGC-----CAAGGCTGGAAGATGT-----374  
 Db 172 yProGlyProGluProGlyArgGlyLeuProAlaAlaGlyAlaArgHisCysArgArgAl 192  
 Qy 375 -----CCAGGTCT-----383  
 Db 192 aAlaProGlyValLeuArgGlnGlyAlaAlaGlyProAlaProProAlaArgSerPr 212  
 Qy 384 -----GGAGCGCGAGCTTGCCTGCAAT 403  
 Db 212 oArgArgAlaGlyGlyGlyIleGlnArgArgGlyArgLeuGlyAsnArgArgLeuProAl 232  
 Qy 404 TTGCGCCCTCGAGGCGCGGACGCTCATCTCTCTCTACTC-----GC 451  
 Db 232 aAlaAlaThrGlyGlnProArgArgGlyProHisTyrSerProValProArgAlaProAl 252  
 Qy 452 TCACGATGATTCACCGTTCCAAACGCTCATCGACGAGCTTCTCGTA-----500  
 Db 252 aAlaProAlaGlyAlaValProArgThrGlyGlyProGlyArgValLeuTyrLeuHi 272  
 Qy 501 -----CCTGTC-----CCAAGCGCTGCTGGCG 526  
 Db 272 sArgGlnProAlaAlaSerArgGlyValProGlyValGlnProArgArgProGly-----290  
 Qy 527 TTGCGGACTTCTACGTGAGCGGCAAGTACGACTGCTCCCTCGC-----569  
 Db 291 ----ArgGlnLeuArgProGlyAspAlaArgProAlaProGlyAlaGlyAspGlyAspAr 309  
 Qy 570 -----CCAGATGCTGCTGCGCGCTTTCTTCTGCGCATCGATCT 610  
 Db 309 gArgGlnGlyAspArgHisProAspArgArgGluGlyLeuArgArgAlaArgArgAspAl 329  
 Qy 611 TCACAT-----CGACAATTCACATCGGCC-----CGAGCGCC 646  
 Db 329 aGlnArgProGlnProAspArgGlnArgGlnLeuArgProGlyGlyAlaArgArgArgPr 349





Db 1056 rgargHisAlaGlnLeuArgLeuValHis-----ArgArgSerGlyAlaAlaGlyH 1074  
Qy 1885 CGCGTCAACATGTACAGCTCTTCTTACATGCCCGCGGAGCGCGCAAGAGGCA 1943  
Db 1074 isArgArgPro-----LeuYrHisProGlyGlnProArgProGlyGlyGln 1089  
RESULT 7  
US-09-252-991A-27068  
; Sequence 27068, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252.991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 27068  
; LENGTH: 638  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-27068  
Alignment Scores:  
Pred. No.: 1.39e-15 Length: 638  
Score: 323.00 Matches: 194  
Percent Similarity: 32.6% Conservative: 36  
Best Local Similarity: 27.5% Mismatches: 239  
Query Match: 8.9% Indels: 236  
DB: 2 Gaps: 38  
US-10-620-914-44 (1-1947) x US-09-252-991A-27068 (1-638)  
Qy 1939 CTTTCTTGGGCGCCCTCCGCGGCGCATGTAGAGGAGCTGTACATGTTGACGCGGTCCA 1880  
Db 66 ProArgGlnArgArgArgAlaProAlaGlyArgAlaAlaSer-----Pro 81  
Qy 1879 TGTAGCCCTAGTGGCGCGGAGATGACGCGCAGCTCGAAGCCCG----- 1835  
Db 82 GlyLysProArg---ArgGlyLeuAlaGlyAlaThrAlaGlyProArgProArgProArg 100  
Qy 1834 -----CCTTCTGATCAGCTCG-----CGTAGGCGGCTGAGGAGCGGCGCC 1787  
Db 101 AlaLysProArgGlyAlaAlaArgGileGlnArgArgProAlaArgArgGlnAla 120  
Qy 1786 AGATGACGATGCCCGCGGCAACCTGCTTGGCCAGGC---ACTCGGCGCAGCTCGTTGG 1730  
Db 121 -----ArgArgArgProAlaGlyProAlaAlaThrGlyArgAlaArgProHisArgGln 138  
Qy 1729 CCAGGGCA----- 1721  
Db 139 ProGlyThrMetValArgLeuProGlyArgGlyAlaAlaProGlyArgArgLeu 158  
Qy 1720 ---TATCACCAGTCCACCTGGTCCATCAGATCACCCTGGTGTAGTGGCGCTTGA 1664  
Db 159 ProHisProAlaAlaAlaGlnProGlyProGlyArgArgArgAlaAlaGileGlyGln 171  
Qy 1663 GCTCTCCATGAAGAAGTGTGTGAGAGCGGTGAGTTGTCCA-----CCAGGCCAC 1613  
Db 172 -----ProGlnProGlyProArgGln 178  
Qy 1612 TCTTGAGGGTGGCAAGCGCCCTCGCGCAGGTAGTGGGCGAGTTCGCGCAGGA--- 1556  
Db 179 Ala---GlyGlyLysArgGlnProGlyProGlyArgArgArgAlaAlaGileGlyGln 197  
Qy 1555 ACTTGGCGGTGAGCAGTGTGTAGTGTAGTGTCTGCTGCGCAGTGTGCGGAGTCTCCG 1496  
Db 198 ProAlaArgAlaGlyArgAlaAspArgGlyLeuAlaAlaAlaArgArgAla----- 214

Qy 1495 CCACGCGCTCCATGTGTGCGCGCATGTAGTTCCTCAATGTGGGATGCGCGTCCGCTTGATCA 1436  
Db 215 ---ArgArgAlaThrLeuGlyGlnArgArgLeuArg----- 225  
Qy 1435 GCGGTACTCTTCCCGCGCAGCG 1376  
Db 226 ---ArgHisAspGlyAlaGlyArgArgArgGlyTrrpArgArgGlnCysArgArgPro 244  
Qy 1375 CCAGGCTCACGAATTTGACGAACACGACACACGCGCTTGGCGCGCTTCTTCCAGGAGT 1316  
Db 245 ArgArgGlyAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 262  
Qy 1315 GGATGAGCATGTTGTGTGCCACAGCGCGCTCTCTCTCATTTGGGCGCGTTCGCGA 1256  
Db 262 ----- 262  
Qy 1255 GCGCTTGAGGCTTTCGCCAGTCCAGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1196  
Db 263 -----ValAlaProHisProArgProArgProArgProArgProArgProArgProArg 276  
Qy 1195 GCTTCCCATCGCGCTGTGTAGTACAGCGCGCTGTGGAAGTACCAGAGCGCTTGACCC 1136  
Db 277 ValArgArgHisArg---ArgGlnThrGlyArgGly----- 287  
Qy 1135 AGAAGTTGTGCTGTTTGGCAGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1076  
Db 288 -----AlaProGlyProGlyHis-----ProGly 295  
Qy 1075 TGCGCGGTGCGCGCTTCCCG 1016  
Db 296 ThrAlaGlyAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 314  
Qy 1015 TGCGCGCGCTTCTTCCAGTCCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 956  
Db 315 TrpAspAspArgProArgPro---AlaHisArgProArgHisArgGlnProAlaAlaAlaThr 333  
Qy 955 CTGCG 899  
Db 334 ProGlyAlaGlyProArgAlaThrGlyAlaGlyAlaAlaGlyLeuGlyArgArgArgLeu 353  
Qy 898 TCAGCAGCGGTGCTTGGGCTGATCTCCATCAGCTCCATCCGCTCGCGCGCGCGCGCGCGCG 839  
Db 354 LeuAlaPro-----HisIleProAlaTyrProHisArgGlyPro--- 366  
Qy 838 ACAGCTGCGGTGTACAGAGGTGGCGGGAACATGGCGCGCGCGCGCGCGCGCGCGCGCGCG 779  
Db 367 -----AlaGlnProGlyArgPheAlaGluProProAlaGlnLeuProArg-----Arg 382  
Qy 778 GCAGGCGCT-----GGCCAACGCTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 743  
Db 383 AlaGlyArgArgAspProProArgArgThrGlyArgArg-----GlnGlnPro 399  
Qy 742 ACAGCTAGTAGGGCG 683  
Db 400 AspProGlyArgGlyGlnGlnArgProAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 417  
Qy 682 CCACACGCGCTCCAGCTTCTGCTCCAGGTAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 623  
Db 418 ArgTyrArgAlaAspProArgArgProArgArgHisGlySerGlyAsp-----ProAla 435  
Qy 622 TGTGATGTGAAGATCGATCGCAGAAAGAAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 563  
Db 436 GlyArgArgArgArgGlnProArgArgArgGlnSerGlnArgArgProGlySerGly 455  
Qy 562 GCA-----GGTCTGCTTGGCGCTCA-----CGTAGA 536  
Db 456 AlaTrpProAlaGlyAspArgArgGluProGlyArgProAlaAlaAlaAlaAlaAlaAlaAla 475  
Qy 535 AGTCGCGCAGCG 476  
Db 476 AlaArgGlnArgProAlaAspArg-----AlaProGlyArg----- 487



QY 475 TGTGAACGGTGGAAATCATCGTCAGCGAGTAGGAGAGGTGATGAGCGTCGCGTGCCT 416  
DB 487 -----  
QY 415 CAGGGGGCGAAATTTGCAAGCGTCGCGCTCCAGACCTGGACATTTCTCCAGCCCTGG 356  
DB 488 -----GlyValGlyAspHisArg-  
QY 355 CCTTCGCTTCTTCTTGGCCACCTCGCACAGCGAGTGGCACAGGTGCGACCACTAGATGG 296  
DB 494 -----GlnArgAlaAspGlyAspProLeuAspArg-  
QY 295 ACTTGAATCTCGCAGGTGATGATCATGACCATCATATGACATTTCTCCCAAGTGCAC 236  
DB 504 -----ArgThrAspGlnProAla-  
QY 235 CACCCAGGTCAACCATGAGTTCGAGCGCTCGCCAGCGGGGCGAG-CAAAGCGAG 179  
DB 515 HisArgGlyArgProArg-ArgArgProGlyProArgLeuArgArgGly 530  
QY 178 CGGCTGGGGCCGTAGAGCTCTCCAGCGGAGCGAGCGTGCATCGC----- 131  
DB 531 ArgArgGlySerProLeuAlaGlyProAlaHisArgArgHisArgGlyAspProAla 550  
QY 130 -----CCTTCTTGTGCGCAACCATATGCGCGAGAAACCGTCACTCCT 83  
DB 551 AlaHisArgProAlaThrGlyArgAlaAlaIleGlyGlyHisAlaGlnProGly 570  
QY 82 TCATGCTGCTGAGCTTCCAGGGAGAGTTCCTTGGTGTAGTCTCCAGGCC 23  
DB 571 ArgAlaCysArgThrHisArgArgThrGlyArgArgGlyGlyArgArgAlaGlyArgSer 590  
QY 22 GGCGCTCAGCACCGC 8  
DB 591 GlyArgArgArgPro 595

## RESULT 8

US-09-252-991A-18296  
; Sequence 18296, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 18296  
; LENGTH: 467  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-18296

## Alignment Scores:

Pred. No.:	3,14e-15	Length:	467
Score:	317.50	Matches:	178
Percent Similarity:	32.7%	Conservative:	26
Best Local Similarity:	28.5%	Mismatches:	198
Query Match:	8.8%	Indels:	222
DB:	2	Gaps:	34

US-10-620-914-44 (1-1947) x US-09-252-991A-18296 (1-467)

QY 1828 GGATCAGCTCGCGTAGGCGGGCTGAGGAGGGCGGCGCAGATGACGATGCGCGCG 1769  
DB 3 GlyCysAlaGlyCysProAlaGlyCysGlyArgArgGly----- 15  
QY 1768 GCGCAACCTGCTTGGCCAGGCACTCGGCGCAGCTGCTTGGCCACGGGCATATCCAGCCAGT 1709

DB 16 -----AsnTrpProAlaArgAlaGlyPheProPro-----ProAlaCys 31  
QY 1708 CCACGTTGTCATCAGAAATCACCTTGGTGTAGTGGCGCTTGTAGCTCCTCCATGAAGA 1649  
DB 32 ArgProGlyThrGlySerSerArgTrpAsnArgAlaGly----- 44  
QY 1648 AGTTGTTGAGAGCGGTGAGTGTCCACCCAGCCACTCTTTGAGGGTGGGGAAGCGCGCT 1589  
DB 45 -----ArgAlaArgGlyCysPro-----GlyCysAlaAlaProPro 56  
QY 1588 CGC-----GCAGTAGTGGGGCAGTTGTCCGCGCAGAACTTCCCGGTGAGGC 1541  
DB 57 AlaGlyValProGlyAlaGlyArg-----SerValProAlaProAlaCysArg- 72  
QY 1540 AGTTGTTAGTAGTAGTCTCTCAATGGGATGCGCTCCGCTTGTATCAGCGCGTACTGCTTGC 1481  
DB 73 -----ArgArgArgHisArgArgArgSerArgProSerSer-----GlyArgProArg 88  
QY 1480 TGCGCGCGATGTAGTCTCAATGGGATGCGCTCCGCTTGTATCAGCGCGTACTGCTTGC 1421  
DB 89 GlnGlyArgArgGlySer-----ProThrAlaAlaLeuArgLeuArg 102  
QY 1420 CGGCGACGCGCGCGCCAGCACAGCGCCTTGTGAAGACGACGACGCTCAGCACT 1361  
DB 103 ProAlaArgArgArgArg-----ArgHisPro----- 111  
QY 1360 TGACGAACAGCCACACCGCGCTTGGCGCGCTTCTTCCAGAGTGGATGAGCATGTTGC 1301  
DB 112 -----AlaTrpArgArgArgSerAlaSerArgProAla- 122  
QY 1300 TGTCACACAGCGCGCTGCTCTCCTCAATTTGGCGCGTGGCGAGCGCTTGCAGCTCT 1241  
DB 123 -----GlyAlaProProPro----- 127  
QY 1240 TGCGCAGTCCAGCACCGCGCGCAGCACTGCGACCCAGC----- 1199  
DB 128 AlaProCysProAlaProArg-----GlyAlaGlyAlaProSerValArgGluArgProAla 146  
QY 1198 ---ACAGCTTGGCCATGCGCGCTGGT-----AGTACAGGC 1166  
DB 147 GlySerAlaProProProProArgProAlaArgArgProProAlaProHisArgThrGly 166  
QY 1165 CGTGTGGAAGTACAGAGCGCTTGGACAGAGTGTGCGTGGTGTGGACAGGAGG 1106  
DB 167 GlySerGly-----TrpArgArgArgGly-----ArgThrArg 178  
QY 1105 GCGCAGCTTCTTCTCTG-----ACAGCTCTCAATGCGCGGTGCGC 1061  
DB 179 ArgProAlaAlaSerArgArgProAlaGlyGlyProAlaGlyArgProGlyGlySerArg 198  
QY 1060 CCTCGCGCAACAGCTGCCACAGCTCTCAAACTCCAGCTGCTGAATGCCACCTTCTTCA 1001  
DB 199 AsnArgArgArgAla----- 203  
QY 1000 GCTCCAGAGCGCGACTGCGCGGGTTGAGTCCACCGACACCACTGCGCGCGCCCT 941  
DB 204 -----GlyCysArgProAlaArgProValAlaArgAla 215  
QY 940 GCACACAGGTTTCAGGGCATTCAGCGCGCGCTAGTACAGGTGAGCACTGCTGCTGG 881  
DB 216 GlyPro---GlyProGlyArgCysArgArgArgArgCysAlaTrpSerProCysPro--- 233  
QY 880 GTTTGTATCTCCATCACCCTCCATATCCGCTCGGGTCTCCAGCACTGCTGCTACAGGA 821  
DB 233 ----- 233  
QY 820 AGTGGGGGGAACATGCGCGCGCTCCAGCGCTCTCTGTCAGGCGGTGGCCACGC 761  
DB 234 -----ThrProArg---AspArgAlaGlyAlaGlyAspArg 244  
QY 760 TGGCGAGCGCGCCAAATCCACAGCTAGTAGGGGG-----CGCGCAGCCAGCGCAGT 710

```

Db 245 HisValProAlaGlyAspAlaArgAlaAlaGlyTrpProThrGlySerArgCysAlaArg 264
Qy 709 AGGGATCGAACCTGGGTGTTCTGCTCCACACGCGCTCCAGCTTCTGCTCCAGGTAGG 650
Db 265 ArgSerGlySerAlaGlyCysAlaArgPro-----ProArgArgSerCysValArg 281
Qy 649 CGCGGCGCTCGGGCCCATGTCATGTTGTCGATGTCGAAGATCGATCGCCAGAGAAC 590
Db 282 TrpAlaAlaArg-----ArgValAlaArgAspSerAlaArgSer 295
Qy 589 GCGCGACCGAGGCGATCTGCGCGAGGCGCAGCTGCTACTTGCCTCAGCTAGAGTCGG 530
Db 296 AspArgProValArgSerAlaLeuArgGlyLeuProAlaAlaArgArgSerArgProArg 315
Qy 529 CAAGCCACAGCCGCTTTGGACAGGTACGAGCAAGCTGTGTCGATGACGCTTGGA 470
Db 316 HisArgArgProGlySer-ValProProValArgThrSerArgAla-----ThrProAr 333
Qy 469 ACGTGCATCATCGT-----GAGCGAGTAGGAGAGGAGTGTGATGAGCGTCG 425
Db 333 GAArgTrpHisAlaArgTrpGlnAlaAlaAlaAlaArgProGly-SerValLeuSerSerA 353
Qy 424 CGGTGCCCTCA-----GCGGCGCGCAAAATGGCAAGCGTGGCCCTCCACGA 380
Db 353 lathProAlaAlaProGlyValArgGlyTyProArgGlyTyAla-ArgAlaProArg 372
Qy 379 CCGTGACATTTCCAGCCCT-----TGGCCTTGGCCTTCTTCTTGGCCA 335
Db 373 ProAlaLeuProGlyArgProAlaAlaArgProAlaArgProArgProCysArgArgPro 392
Qy 334 CTTGCGACAGGAGTGCACAGTCGACCA-----CGTAGATGACT 293
Db 393 AlaAspArgSerProGlyGlnGlyArgProProGlySerSerAlaThrArgArgTrpArg 412
Qy 292 TGAATTCGCCAGTCGATGATTAATCAGCCATCATCGACATCTCCCGCAGTCCACAC 233
Db 413 -----ProCysHisArg 416
Qy 232 CCAGTCAACCCAGATGAGTTCGAGCGCTCGGCCAGC-----GGGCAG 188
Db 417 ArgAlaGluProAlaAlaProAlaGlyTyProGlyAlaAlaAlaArgSerAlaThr 436
Qy 187 CAAGGACGCGCTGGGCGCGTAGAAGC---TCTCCAGGCGAGCAGCGTATCATCGC 131
Db 437 ArgArgThrHisAlaGlyArgArgGlySerAlaThrProSerGlyArgArgAspArg 456
Qy 130 CTTTCT 125
Db 457 ProAla 458

```

## RESULT 9

```

US-09-252-991A-31760
; Sequence 31760, Application US/09252991A
; Patent No. 5551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31760
; LENGTH: 1706
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31760

```

## Alignment Scores:

```

Pred. No.: 6,27e-15 Length: 1706
Score: 316.50 Matches: 193
Percent Similarity: 33.1% Conservative: 44
Best Local Similarity: 26.9% Mismatches: 249
Query Match: 8.7% Indels: 231
DB: 2 Gaps: 36

US-10-620-914-44 (1-1947) x US-09-252-991A-31760 (1-1706)
Qy 1846 CGTCGAAGCCCGGCTTCTGATCAGCTCGCGGTAGGCGCGGCTGAGGAGCGGAGCGCC 1787
Db 276 ArgArgGlnProAlaGlyAlaThrAlaArgLeuArgHisArgProGly---HisProAla 294
Qy 1786 AGATGACGATCGCCCGCGCAACCTGCTTGGCCAGGCACTCGGCCAGCTCGTTGGCCA 1727
Db 295 ArgLeuArgArgGlnProAlaArgLeuArgSerGlyGluArgProAlaGluSerPro 314
Qy 1726 CGGCGCATATCCAGCCAGTCCACGTGT----- 1700
Db 315 ArgArgThrProGlyGluLeuArgGlyArgProAspAlaLeuProArgLeuAlaGly 334
Qy 1699 ---CCATCAGATCACCT-----TGTGTAGGTGCGCGCTTTGAGCTCT---CCATGA 1652
Db 335 LysProGlyArgTyProAlaGluPheAlaArgGlyAlaAlaProAlaProGlyPro--- 353
Qy 1651 AGAAGTTGTGGAGAGCGTCAGTGTTCACACGCGCCTCTTTGAGGG-----TGG 1601
Db 354 -----LeuAlaAlaValProArgArgHisAlaArgAlaAlaGluProTrp 369
Qy 1600 CGAAGCGCGCTCGCGCAGTAGTGGGCGAGTGTTCGCGCAGGACTTCCGCGTGGAGC 1541
Db 370 ArgArgProAlaGlyGlyArg-----AlaAspArgThrAlaThr----- 383
Qy 1540 AGTTGTAGTAGAAGTAGTTCTGCTGCGCACGTCGAGTTCTCGCGCAGCGCTCCATGG 1481
Db 384 AlaProAlaArgProArgThrAlaCysAlaAlaAlaValAlaGlyProArgHis----- 401
Qy 1480 TGCGCGCATGTAGTTCTCAATGGGATGCCGCTCGCGCTTGATCAGCGCTACTGTGTC 1421
Db 401 ----- 401
Qy 1420 CCGGCACGCGCGCGCGAACCAGCACGCGCTTGTGTGAAGACAGCAGCTCACGA--- 1364
Db 402 ---AlaArgGlnGluProArgThrLeuProGlyProArgProGlyAlaArgGly 420
Qy 1363 -----ACTTGACAAACAGCACAGCGCTTGGGCGCTTCTTCACGA 1319
Db 421 ThrHisLeuAlaGlySerAlaGlnProAlaAlaProAlaAla---AlaArgProPro--- 438
Qy 1318 AGTGGATGAGCATGTTGCTGTCCACAGACGCGCTGCTCTCTCATTTGTGGGCGGTGG 1259
Db 439 -----AlaAlaLeuProAlaAlaArgAlaArg--- 446
Qy 1258 CGAGGCGCTTGACGCTTGTGCCCGCTCCAGTCCAGCACCA-----CGGCCAGCTACGA 1208
Db 447 ArgGlyArgArgGlnProProGlyArgProArgArgLeuArgProGly---Ala 465
Qy 1207 GCACCCAGACAGCTTGGCCCATGCCCGCTGTGTAGTACAGCGCTGTGGAAGTACCAGA 1148
Db 466 ProValAlaAlaAlaAlaPro-----AlaThrGlyGlyAlaAlaThrGluPro 481
Qy 1147 GCGCTTGGACCAAGATTTGTGCTGTTTGCACAGGAAGGCGCGCAGCTTCTTCTCGT 1088
Db 482 GlyAla-----GlyAspSerArgHisArgCysArg 491
Qy 1087 ACAGCTCTCAATGCGGGGTGACGCGCTTCGCGGAACAGCTGCCACAGCTCTCAAACT 1028
Db 492 SerGlyThrAspArgProGly---ArgProArgGlnProAlaAlaAlaArgAspProGly 510
Qy 1027 CCAGCTGCTGAATGGCCACCTTCTTCAGCTCCAGAAAGCGCGCAGTGGCGGGTTGCAGT 968
Db 511 ProGlySer---TrpProGlyProSerAlaAlaArgAlaAlaArgGly---Asp 528

```

```
Qy 967 CCA-----CGGACA 959
Db 529 ProHisArgArgGluThrAlaLeuArgSerAlaAlaArgArgProArgArgAlaAsp 548
Qy 958 CCACCTGGCGGCGCCCTGCACACAGAGTTTCAGGGCATTTCAGCCCGCGCTAGTCAGGG 899
Db 549 ProProGlyArgThrProAlaGlnProGlyGlnThrGluGlyAlaArgArg----- 565
Qy 898 TCAGCACCCTGCTCTGGGGTTGATCTCCATCATCTCCATATCGCGGCTCGGGTCTCC 839
Db 565 ----- 565
Qy 838 ACGACTCGGTACAGGAAGTGGCGGGAACATGGCGCGCTCCACGGCTCTCTCGT 779
Db 566 -----LeuAlaArgArgThrAlaGlyThr-----AlaAlaProArg 578
Qy 778 GCAGGGCGTGGCCCAACCTGGCGAGCGGCGCAATCCACAGTATAGGGGGCGCGCAGCC 719
Db 579 HisArgArgGlnProGluGlnProAlaGlyProArgProAlaArgGlyAlaThrAla 598
Qy 718 ACGGCACGT-----AGGGATCGAACCTGGG 692
Db 599 ThrSerArgLeuProAlaLeuProTrpProArgArgProAlaArgSerHisArgProGly 618
Qy 691 TGTCTCTCCACACGGCTCCAGCTTCTCTCCAGTAGGCGCGCTCGGGCGCA 632
Db 619 IleProGlyProAla-----AlaProAlaProGlyThrThrArgArgAspSerArgArg 637
Qy 631 TGTCAATGTTGTCGATGTCGAAGATCGATGCCAGAAAGAACGGCGGACCGAGGCA--- 575
Db 638 -----ThrArgSerHisGlyAlaProHisAlaGln 647
Qy 574 -----TCTGGCGCAGGGCAGGTGCTACTTGGCGCTCAGTAGAAGTCGGCAACGC--- 524
Db 648 ProArgSerAlaAlaGlyLeuGluAlaAlaProGlyThrArgGlnSerGlyThrAlaVal 667
Qy 523 CCACGAGCGCTCTGGGACAGGTACGAGCAAGCTGTCGATGACGTGTGGAAACGGTG 464
Db 668 GlnProGlyArgSerGlnGlnGlyArgGlnProArgArgSerArgCysAlaAspAla 687
Qy 463 GAATCATCTGTCGAGCGAGTAGGAGGTGTAGCGCTCGCGTGCCCT---CAGGGGGCG 407
Db 688 ArgAlaGluProArgThrAlaArgGlnGlyHisArgArgAlaProGlyGlnProGln 707
Qy 406 CAATTTGGCAGCGTGGCTCCAGCAGCTGGACATCTTCCAGCCCTTGGCCTTCGCT 347
Db 708 AlaMetGlyArgArgHisProPro-----AspArgGln 718
Qy 346 TCTTCTTGGCCACCTCGCACAGCGAGTGGCACAGCTGCACACCTAGATGAGCTTGA 287
Db 719 AlaLeuAlaProAlaAlaArgValHisArgAlaGlyArgArgThrArgProAspGluPro 738
Qy 286 TCGCCAGCTCGATGTAATCAGCCATCATATCGACATCTTCCCCAGTCGCCACCCAGGT 227
Db 739 AlaProGlyLeuAlaAspGluPro-----ArgProGly 749
Qy 226 CAACCCAGATAGGTTTCAGCGCTCGGCCA----- 197
Db 750 GlnProArgAlaAlaAlaGlnGlyArgAlaValArgHisGlyAspLeuArgSerLeu 769
Qy 196 -----GGCGGCGAG 188
Db 770 AlaAspAlaGlyValArgProAlaAspProLeuSerArgAlaGlyHisGlyGlyGln 789
Qy 187 CAAGG-----CAGCGCGCTGGGGCCCGCTAGA 161
Db 790 ArgArgArgGluAlaAspAlaAspArgValLeuLeuGlnProArgGlyLysArg 809
Qy 160 AGCTCTCCAGGAGCAGCGTATCATCGCCCTTCTTCTCGCCGACCCATATATGGCGCA 101
Db 810 GlyArgThrValArgArgSerAlaArgArgGlyArgArgArg-----GlyAla 826
Qy 100 GAACGGTCA-----GGTCATCTTCTATGCTGTGCTGAGCTTCTCCAGGGAGAGT 47
```

```
Db 827 AlaArgSerLeuArgGlyHisLeuGluProProArgAspGlnGlyLeuProGlyProPro 846
Qy 46 TCTTCTTGGTGTAGCTCGCAGCGCGCGCTCAC-----GACCGGACCCCA 2
Db 847 AlaThrGlyAlaGlnArgProAlaGlnHisHisProAlaAspProLeuPro 863

RESULT 10
US-09-252-991A-27341
; Sequence 27341, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27341
; LENGTH: 1073
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (803)
; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
US-09-252-991A-27341
```

```
Alignment Scores:
Pred. No.: 1.58e-14 Length: 1073
Score: 310.00 Matches: 243
Percent Similarity: 28.1% Conservative: 41
Best Local Similarity: 24.0% Mismatches: 288
Query Match: 8.6% Indels: 440
DB: 2 Gaps: 47

US-10-620-914-44 (1-1947) x US-09-252-991A-27341 (1-1073)
Qy 18 CGGCGGCGCTCGCAGCTACAC-----CAAGAAGAACTTCTCTCGGAGAGAGCTCAA 68
Db 92 ArgCysSerCysThrLeuGlyAlaSerIleArgGluGluArgGlnProAspGluArgLys 111
Qy 69 GCTCAGCAGCATGAAGGATGACCTGACCCGTTCTGCGCCATATGTGTTCGGCAGCAAGAA 128
Db 112 AsnAlaArgHis-----ValProArgArg 120
Qy 129 GGGCGATGATCAGCTGCTCTCGCTGGAGAGCTTCTACGGGCGCCCGCTGCTTGC 188
Db 121 ArgArgProGlyProCysArgProGlyProLeuHisAlaArgLeuProArgValArgArg 140
Qy 189 TGCCCGCTGCGCGAGCGCTCGAACCTCATCTGGGTTGACCTGCGGTGCTGCGCACTGGGGA 248
Db 141 ProProArgGlnArgLeuAlaGlu----- 148
Qy 249 GAATGTCGATATGATGGCTGATTATCATCGACCTCGCGAAGTTCAAGTCCATCTACGTGGT 308
Db 149 -----HisArgArgArgArgArgArgLeuLeuArgHis 160
Qy 309 CGACCTGCGCAGCTCGCTGCGAGGTGGCCAAAG----- 344
Db 161 ArgArgGlnProLeuProArgArgGlyGlyArgHisValHisGlnHisArgProGly 180
Qy 345 -----GAAGCGAAGGCGCAAGGCTGGAAGATGTCCAGGT----- 380
Db 181 AlaArgGlyAsnGlySerHisArgGlyArgAlaAspProProAlaGlyLeuPheGlnSer 200
Qy 381 -----CGTGGAGGC 389
```



QY 1713 GCTGGATATGCCGTGGCCACGAGCTGGCGAGTGCCT-----GGCCAGGAGGCT 1763  
Db GlyAlaValArgArgGlyProPheAlaProArgHisAlaAspArgLeuGlyGlyAlaGly 925  
QY 1764 TGGCGCGGCGGCGCATCGT----- 1781  
Db LeuGlyGlyGluHisArgProAlaAlaLeuArgProGlyGlyAlaHisCysArgGly 945  
QY 1782 -----CATCTGGCGCTCCGCTCCCT 1802  
Db ThrValLeuAspProAlaValAspLeuArgHisGlyPheHisAspAlaLeuCysAlaAsp 965  
QY 1803 CAGCCGCCCTACCGCGAGCTGATCCAGAA----- 1832  
Db ValProGlyLeuArgProProAlaProArgArgProAlaSerThrAlaLeuProAspAla 985  
QY 1833 -----GGCGGGCTTCGACGTGGCGCTG----- 1853  
Db AlaGlyArgProGlyGlnSerLeuGlyAlaValArgAlaAlaCysAlaGlyArg 1005  
QY 1854 -----CATCCGCGCGCCACTCAGGCTACATGACCGCTCAACAT 1895  
Db HisLeuProValArgHisProArgArgAlaAspGlyLeuGlyLeuArgArgGlnAsp 1025  
QY 1896 GTACAGCTCCTTACATGCGCGCGCGAGGCGGC 1931  
Db 1026 ArgArgArgArgGlyAlaGlyProGlyArgGlyArg 1037

## RESULT 11

US-09-252-991A-23774  
; Sequence 23774, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 23774  
; LENGTH: 798  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-23774

## Alignment Scores:

Pred. No.: 1.66e-14 Length: 798  
Score: 309.00 Matches: 213  
Percent Similarity: 30.4% Conservative: 33  
Best Local Similarity: 26.3% Mismatches: 255  
Query Match: 8.5% Indels: 309  
DB: 2 Gaps: 47

US-10-620-914-44 (1-1947) x US-09-252-991A-23774 (1-798)

QY 1933 TGGCGCCCTTCGCGCGGG-----CCATGTAGAGGAGC 1901  
Db TrpProMetAlaGlyGlySerMetAlaCysValAlaArgTrpThrProCysArg----- 33  
QY 1900 TGTATCATGTTGACCGGGTCCATGTAGTCCCTGAGTGGCGCGCGGATGACGCGACGTGCA 1841  
Db 34 -----ArgGlySerCys-----ArgTrpLeuAlaIleCysArgLeuArgSer 47  
QY 1840 ACCCGCGCTTGTGATCATCTCGGCGTAGGGCGGCGCTG-----AGG 1800  
Db ValProArgSerGlyArgAlaSerArgSerThrTrpCysThrAlaProProVal 67  
QY 1799 GAGCGGAGCGCCAGATGACGATGCGCGCGCGCAAC-----TGCTTGCGC 1752

Db 68 SerThrThrArgLeuAlaThrValProProThrSerThrAlaCysAlaMetCys-TripAl 87  
QY 1751 AGGCACTCGCGCCAGCT----- 1736  
Db 87 aGlyCysArgProAlaAlaSerGlyProGlyAlaCysCysSerSerProAlaProGlyse 107  
QY 1735 -----CGTTGGCCACCGGCATA 1719  
Db 107 rThrArgArgProMetAlaAlaGlySerThrArgSerArgArgArgPheProArgSerIl 127  
QY 1718 TCCA-----GCCAGTCCACGTGGTCCATCAGATCACTTGGTGT----- 1679  
Db 127 eProAspGlySerCysSerThrProSerArgSerArgTrpThrAlaGlyTyrArgLeuPr 147  
QY 1678 -AGGTGGCGCTTTGACCTCTCCATCAGAGTGTGGTGGAGCGGTCCAGCTTCCACC 1620  
Db 147 oAlaCysAlaTrpProAlaSerThrValArgGly-----GlySerGlyCysSerTh 164  
QY 1619 ACGC-----CACTCTTTGAGGTGG-----CGAAGGCGC----- 1592  
Db 164 rArgCysValArgGlyThrGlyTrpSerAlaSerArgProCysMetProThrGlySerTh 184  
QY 1591 -----CCTCGCGCAGGTAGGTGGCGAGTGTGTGGCGAGGAACCTTCCCGGTGAGGAG 1539  
Db 184 rProThrMetArgProGlyCysTrpProSerCys-----CysAlaProMetPr 200  
QY 1538 TTGTAGTAGAAGTAGTTCCTGCTTGGCGCAGTTCGCGCAGCGCTCCATCG----- 1481  
Db 200 oAlaAlaArgArgTrpArgThrAlaThrSerValSerThr-----ThrArgProTrpAr 218  
QY 1480 -TGCGCGCGA-----TGTAGTTCTCAATGG-----GGATGCGCTCGCCTTGATC 1437  
Db 218 gCysThrArgTrpSerThrCysAlaSerAlaTrpAlaSerAlaSerAlaGlyProThrAs 238  
QY 1436 AGCCCGTACTGCTTGGCGCGCACGCCCGCGCGCAACACACGACGCGCTTGTGAGAGC 1377  
Db 238 nileArgTyrAlaAlaProAlaAlaSerValAlaAlaThrAlaAlaProGly---ArgSe 257  
QY 1376 ACAGGCTCAGCACTTGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1317  
Db 257 rAlaGlyCysArg-----AlaIleProAlaThr---AlaArgAlaThrArg- 271  
QY 1316 TGGATGAGCATGTTGCTTCCACAGACGCGGCTCTCTCCATTG-----TGGGC 1266  
Db 272 -----ArgCysSerGlyThrProAspGlyCysGlnProValValGlyAlaSerLeuAl 289  
QY 1265 GCGTTGGCGAGCGCTTGACGCTTGTGCCAGTCCAGCACGACGACGACGACGACGACGACG 1206  
Db 289 aAlaTrpArgGlyAla-----ValProCysGlyLeuLeuArgProGlyValArgPr 306  
QY 1205 ACCCAGC----- 1199  
Db 306 oProGlnProPheGlyLeuArgHisArgLeuAspProProArgArgArgArgLeuProVa 326  
QY 1198 -----ACAGCTTGGCCATGCCCGCCCTGGT-----AGTACAGCGCGCTGC 1161  
Db 326 lProArgAlaAlaProCysAlaAlaGlyArgHisProValAlaAspGluProGlyArgAl 346  
QY 1160 TGAAGTACAGAGCGCTTGGACGAGATGTTGTGGCTGG-----TTTGCACAGC 1110  
Db 346 a---ThrTyrArgAlaArgArgGileArgAlaAlaGlyLeuProAspAlaLeuCysArgG 365  
QY 1109 AAGGCGCGCAGCTTCTTCTCGTACAGCTCTCAATCGCGGGTGCACGCCCTCGCGCAAC 1050  
Db 365 yAsnAlaProAlaArgPro---AlaArgProGlnThrValAlaAlaArgLeuArgTh 384  
QY 1049 AGCTGCCACAGCTCTCAAACTCCAGCTGTCTGAATGGCCACCTTCTTCAGCTCCGAAGC 990  
Db 384 rGluProGlyArg---ArgArgGlnArgGlyGlyTrpProGlyValAlaGlyArgGlyPh 403  
QY 989 GCCGACTCGCGGGGTTCAGTCCACCGACACCACT----- 953  
Db 989 GCCGACTCGCGGGGTTCAGTCCACCGACACCACT----- 953

Db 403 eProAlaGlnArgArgIleGlyValGlyLysProProAlaGlyLeuAlaArgThrGlyLe 423  
Qy 952 -----GGCGGCGCCCTCGCAGCAGGAGTTTCAGGCGCATTCGACGCGCCCTAGTCAGG 900  
Db 423 uArgAlaSerArgArgLeuAlaProGlyGlyArgGlyAlaProArgArgArg----- 440  
Qy 899 GTCAGCAGCGGTCTTGGGGTTGATCTCCATCATCTCCATATCCGCTCGGGTCTCTCC 840  
Db 441 -----AsnHisArgThrProAlaArgLeuProG1 450  
Qy 839 CACGACTCGGTGTACAGGAGGTGGCG-----GGACATGGCGCGCTCCACG 789  
Db 450 ySerAlaCysTyArgGlyAlaArgSerGlyProThrGlyAlaValAlaAlaProPr 470  
Qy 788 CGCTCTCTGTGCAGGCGGTGGCCAACTGCGGCGGCGCAATCCACAGTAGTAGGG 729  
Db 470 oAlaPro-----ProAspArgG1 476  
Qy 728 GCGCGCAGCCAGCGCAGTAGGGATCGAACCTGGGTGTTCTGCTCCACACGCGCTCC 669  
Db 476 yPheProAlaCysLeuArgArgAlaAlaAlaTyLeuProThrAlaMetProGlyAlaPr 496  
Qy 668 AGCTTCTCTCCA-----GCTAGCGCGCGCTCGGCGCGATGTCATGTTGTCGATG 615  
Db 496 oGlyGlnAlaProValAlaGlyArgThrAlaAlaGlyGlyThrArgProAlaProArgLe 516  
Qy 614 TCGAAGATCGATCCGCCAGA----- 596  
Db 516 uLeuArgProGlyAlaLeuLeuHisAspLeuProAlaPheHisArgGlyHisSerGlyAl 536  
Qy 595 -----AGAACGCGCGCGACCGAGCATCTGG 570  
Db 536 aValProAlaGlnArgGlnAspLeuSerPheGlnTyArgAlaArgAlaAlaG1 556  
Qy 569 CGCAGGCGCAGGTCTACTTCCGCTCA-----CGTAGAGTCCGCA 528  
Db 556 nThrGlyGlyLeuArgValSerArgGlyLeuHisValArgValArgProCysArgG1 576  
Qy 527 ACGCCCA-----CCAGCGCTCTTGGGACAGGTACGAGCAAGCCTGTCGATGACG 477  
Db 576 nTyProPheArgArgProAlaValAlaGlyAlaGlyLeu----- 589  
Qy 476 TTGTGGAAACGTGGATCATCTGTCAGCGAGTAGGAGAGGTGATGAGCGTGGCGGTGCC 417  
Db 589 ----- 589  
Qy 416 TCAGGGGCGCAATTCGCAAGCGTGGCGCTCCAGACCTGGACATTCCTCCAGCCCTTG 357  
Db 590 -----LeuProValGlyAlaArgArgPro----- 597  
Qy 356 GCCTTCGCTTCTTCTTGGCCACCTCGCACAGCGAGTGGCACAGGTGACACAGTAGATG 297  
Db 598 -----AlaAlaArgPro----- 601  
Qy 296 GACTTGAACCTTCGCGAGTGCATGTATATCCACCATCATATTCGCAATTCCTCCAGTGCCA 237  
Db 602 -----AlaProGly-----ArgArgGlyGlyArgArgHi 611  
Qy 236 CCACCCAGGTCA-----CCAGATGAGGTTCGAGCGCTCGCGCCAGCGGGCGAGCA----- 185  
Db 611 sArgProGlyGlnProProGluHisAlaAlaGlyAlaArgProValGlyGlnArgAlaG1 631  
Qy 184 -----AGCAGCGG-----CTGGGGCGCGTAGAGC----- 158  
Db 631 yLeuGlyAlaGlnArgLeuAlaAlaAlaGlyAlaArgArgThrValLeuProLeuAl 651  
Qy 157 -----TCTCCAGGAGCAGCGTGCATATCCGCTTCTGCTCGCGACCAACATATGCGCG 102  
Db 651 aGlyGlyProGly-----AlaAlaCysAlaGluAlaThrGlyAsnAl 665  
Qy 101 AGAACGTGAGTGCATCTTCATGCTGCTGAGCTTTCAGGCTTTCAGGAGAGATGTTCTTC 42  
Db 665 aGlyAlaAlaArgArgPro-----ValProArgArgLeuSerAl 678

Qy 41 TTGGTGTAGCTCGCAGCGCGGCC 19  
Db 678 a-ArgAlaAlaArgArgProAla 685

## RESULT 12

US-09-252-991A-17508  
; Sequence 17508, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 17508  
; LENGTH: 1064  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-17508

## Alignment Scores:

Pred. No.: 2,21e-14 Length: 1064  
Score: 308.00 Matches: 215  
Percent Similarity: 33.9% Conservative: 43  
Best Local Similarity: 28.2% Mismatches: 267  
Query Match: 8.5% Indels: 237  
DB: 2 Gaps: 44

US-10-620-914-44 (1-1947) x US-09-252-991A-17508 (1-1064)

Qy 149 GCCTGGAGAGCTTCTAGCGGCCCGCCGCTGCTTGTGTCGCC----- 193  
Db 9 AlaTTPSerAlaAlaAlaTTPProTTPProAlaSerSerProCysSerSerCys 28  
Qy 194 ---GCCTGGCGGAGC-----GCTCGAACCTCATCTGGGTGACCTGGGTG 235  
Db 29 CysAlaTTPProSerAlaProProGlyProAlaThrAlaCysSerCysCysSer 48  
Qy 236 GTGCGACTGGGAGAGTGTGATATGATGCTGATGCTGATGCTGACCTGCGAGTCAAGT 295  
Db 49 AlaAlaCysSer-----CysSerAlaTTP---AlaCysSer 59  
Qy 296 CCAATCTACGTGTGCGACCTGTGCCACTGCTGCGAGGTGGCCAGAGGCGAAGG 355  
Db 60 AlaSerThrSerGlyAlaCysThrAlaThrCysAlaArg----- 72  
Qy 356 CCAAGGGCTGGAAGATGTCAGGTG-----TGAGGCGCGAGCTGTCCTTCGCG 409  
Db 73 -----GlyArgAlaSerSerSerAlaTTP-----SerAlaGlnArg 85  
Qy 410 CCCTCGAGGCGCAGCGAGCTCATCA-----CCTTCTCTACTGCTCAGCATGATTC 463  
Db 86 ProProArgPheProArgProCysSerGlyProValSerProLeuHisProAlaSerLeu 105  
Qy 464 CACCGTTCACACAGCTCATCGACAGCTTGTCTACCTGTCTCCCAAGA----- 512  
Db 106 ProArgHisAspLeuGluSerArgArgLeuArg-LeuProArgHisArgLeuHisArgTy 125  
Qy 513 -CGGCTGTGGCGGTGGCGACTTCTAGTCAGCGGCAAGTACGACCTCCCTCGCGCC 571  
Db 125 xArgSerProAlaGlnCysArgLeuArgAspArgArgLeuHisProCysArgArgPr 145  
Qy 572 AGATGCCCTGGTGGCGCGCTTCTTCTGGCGATCGATCTTCGACATCGACAACTTGACA 631  
Db 145 oThrGlyLysHisLeuLeuArgLeuGlyArgThrProLeuCysArgAlaArgHis----- 163



Qy	632	TCGGCCCCGAGCCCGCGCTACCTCGAGCAGAGAGCTGAGCGCGTGTGGAGCGCAGACA	691
Db	164	-----SerAlaAlaArgAlaArgGlyArgGluProSerAlaValAlaGlyAla---Hi	180
Qy	692	CCGAGGGTTCGATCCCTACGTCCGTGCGCGCCCTTACTACGTGTGATTTGGCC	751
Db	180	sProProThrAlaPro-----GlyLeuProValLeuLeuLeuLeuProPr	195
Qy	752	GCCTGCCAGCTGTGGCCACGCCCTCAGCA	787
Db	195	oProAlaAArgArg-----ArgAlaAlaArgLeuArgThrArgArgLeuGlnProAl	213
Qy	788	GCTGTGAGCGCGCGCCATGTTCCGCCACCTTCTGTACAGCGACTGCTGGAGGACC	847
Db	213	aArgPheAlaAlaAlaLeuProArgThrSerProGlyGluLeuGlyAlaGlycInAr	233
Qy	848	CCGAGCGGNATGGAGGTGAT-----GGAGATCAACCCCAAGGACACGGTCTGA	898
Db	233	gArgAsnAlaAspArgGlyAspProAlaSerHisAspArgAlaArgArgHisArgAlaAs	253
Qy	899	CCCTGACTAGCGCGGTGCAATCCCTGAACCTGTGTGTCAGGGGGCGGCCAGGTGG	958
Db	253	p-----ProArgProAlaGlyArgArgHisArgProArgGl	265
Qy	959	TGTCGTGTGACTGCACACCCGCGCAGTCGCGCTCTCTGGA-----GCTGAAGAAGGTGG	1012
Db	265	yThrArgPro--GluProAlaArgGlnAlaAlaGlnGlyArgArgCysProAlaAlaAr	284
Qy	1013	CAATTACAGACGTGGAGTTTGAGACGTGTGGCAGCTGTTCGGCAGGGCGCT	1064
Db	284	gGlnProAlaAlaAlaArgProArgArgAlaAlaGlySerArgAlaGlyArgGluProGl	304
Qy	1065	-----GCACCGCGCATTTGAGAGCTGTACGAGAAGAAGCTGGCGCCTCTCTCT	1114
Db	304	yGlnProLeuArgProAlaHisProGly-----	313
Qy	1115	CGCAACACGCCACCACTTCTGTGTCCAGCGCCTCTGTACTT-----	1157
Db	314	---GlyArgProAlaArgLeuAlaGlnAlaGlyThrAlaValValArgProGlyAlaAr	332
Qy	1158	-----CGACACGGCT-----GTACTACAGCGCGCATGGCAAGCT-----	1196
Db	332	gGlyAspProAlaLeuProTrpArgLeuLeuProGlyArgArgThrGluAlaAspArgIue	352
Qy	1196	-----	1196
Db	352	uGluArgArgGlyArgArgGlnProArgSerArgAlaGlyGlnArgThrValLeuAr	372
Qy	1197	-----GTGCTGGGTGTGCA-----	1211
Db	372	gProAlaAlaAspArgLeuArgArgGlyLeuAlaGlyAlaAlaLeuArgProAlaArgAr	392
Qy	1212	-----GTGCTCGCGCT-----GGTGTGGACTGGGCAAGACC	1246
Db	392	gAlaArgProValProGlyArgHisAlaThrGlyHisArgAlaGlyProGlyArgGlyAr	412
Qy	1247	TCAGCGCCT-----CGCAACCGCGCCACAATGGAGAGCAGCGCGCTC	1291
Db	412	g--AlaProAlaTrpArgGlyValArgProAlaAlaProHisAlaGlyAlaAspProAr	431
Qy	1292	TGTGGGACAGCAACATGCTCATCCACTTCGTGAAGAACGGCGCCCAAGCCGCTGGTGGC	1351
Db	431	gGlyGlnArgLeuHisArgGlnPro-----ProValArgThrProAlaAlaArgArgSe	449
Qy	1352	TGTTCGTCAAGTTTGTGAGCCTGTGTCTTTCAA-----CAAGCCGTGC	1396
Db	449	rLeuArgGlyProArgHisGlyHisArgLeuArgArgHisArgThrAlaGlnGlyArgPr	469
Qy	1397	TGTG-----GTTTGGCGCGCGCT-----GC	1417
Db	469	oAlaPheProLeuArgArgArgArgHisArgHisProPheGlyValAlaArgIleProCy	489
Qy	1418	CGGGCAACGACGTACGCGCTGTATCAAGCGCGGACCGCATCCCATTTGAACTACATCGCGC	1477

Db	489	sGluGluMetArgAspProAlaalaGlyGlyHisArgHis-----AlaHisArgVa	507
Qy	1478	GCAC-----CATGACGGCGTGGCGAGAACTCGCAGCTGGCGCAACGACGAACT	1525
Db	507	lHisAlaGlnProAlaalaGlyValArgThrGlyLeuArgGlyLysProAlaAspArgPr	527
Qy	1526	ACTTCTACTACAACCTGCCTCAC---CGGCAAGTTCTCT-----CGCGCAACTGCCCCA	1576
Db	527	oLeuLeuArgGluileArgGlnThrArgGlyValProPheHisLeuArgGlyLeuArgHl	547
Qy	1577	CCTACCTGCGGAGGCGGCTTCGCCACCTCAAGAGTGGCGTGGTGGACAACCTGACCG	1636
Db	547	sValProGlyProGlyLeuArgArgProLeuGlu---ProGlyGlyArgAlaAspGl	566
Qy	1637	TCTCCACCAACTTTCTTCATCGAGAGCTCAAGCGCG-----1673	
Db	566	nGlnAlaAlaLeuAspLeuGlyValGlnAlaValAlaAaArgProGlyAspLeuGlyLe	586
Qy	1674	-----CACCTACAC-----CAAGGTGATTCTGA	1696
Db	586	uArgProAlaGlyProAlaLeuHisProValProSerValGlnLeuAspGlyProAlaPr	606
Qy	1697	TGGACCACTGGA-----CTGGCTGGA-----TATGC	1723
Db	606	oGlyProProGlyPheGlyAlaAspArgGlnLeuAlaGlyAspHisProAlaHisProAl	626
Qy	1724	CCGTGGGCCAACGAGCTGCGCGAGTGCCTGCCCAAGCAGTTGGCC-----1769	
Db	626	aProGlyArgArgHisAlaAspProProGlyArgArgArgAlaGluAlaLeuLeuHi	646
Qy	1770	-----GGCGCGCATCGTCATCTGGCGCTCCGCTC-----CCTCAGCCCGCCCT	1813
Db	646	sArgArgArgArgHisArgGlyProArgAlaAspHisArgGlnProArgProLe	666
Qy	1814	ACGCGGAGCTGATCCAGAGCGGGCTT-----CGACGTGCGCTGATCGCGCGGCCA	1867
Db	666	uArgArgAlaAspArgGlnHisArgGlnProGlyGlnArgGlyGluHisProProAlaAr	686
Qy	1868	CTCAGGCGCTACATGACCGCGCTCAACATGTACAGCTCTCTTCATCGCGCGCGGAAGG	1927
Db	686	gArgGlyThrAlaAlaPro-----ValArgGlyProProAlaAla	699
Qy	1928	CGCG 1931	
Db	699	aArg 700	

RESULT 13  
 US-09-252-991A-20408  
 ; Sequence 20408, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252.991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 20408  
 ; LENGTH: 957  
 ; TYPE: PR1  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-20408

Alignment Scores:	2,31e-14	Length:	957
Pred. No.:	307.50	Matches:	224
Score:	34.6%	Conservative:	32



Best Local Similarity: 30.3% Mismatches: 280  
 Query Match: 8.5% Indels: 204  
 DB: 2 Gaps: 40

US-10-620-914-44 (1-1947) x US-09-252-991A-20408 (1-957)

QY 29 CGAGTACACCAAGAGAACTCTCCCTGGAGAGCTCAAGCTCAGCAGCATGAAGGATG 88  
 DB 128 ArgArgProAlaArgThrArgPro-----136  
 QY 89 ACCTGACCGTTCTGCCCATATGGTTCTGGCAGCAAGAGGGCGATGATCACGCTGCTC 148  
 DB 137 -----LeuProGlyProAlaAlaSerProAla-----ThrGly 147  
 QY 149 GCCTGGAGAGTTCTACGGCCCC-----AGCGCGCTGCTTTGCTGCTCCGCC 196  
 DB 148 AlaGlyArgLeuProProGlyProGlyArgGlyAlaArgProLeuProAlaAlaProPro 167  
 QY 197 TGGCGGAGCGTCA-----ACCTCATCTGGGTTGACC 229  
 DB 168 GlnSerProAlaArgGlyProGlyAlaAlaGlyHisGlnProAlaAlaAlaGlyLeu--- 186  
 QY 230 TGGGTGGTGGCACTGGGGAGAATGCGATATGATGGCTGATTACATCAGCTGCGGAAGT 289  
 DB 187 -----AlaProGlyAlaProAlaProValAlaAlaAlaArgThrGlnArgThr 203  
 QY 290 TCAAGTCCATCTACGTGGTGCACCTGTGCCACTCGCTGTGGAGGTGGCCCAAGAGGAGG 349  
 DB 204 GlyProAlaProAspArgArgGlyGluArgGlyAlaAlaAspGlyArgGlnArgProArgThr 223  
 QY 350 CGAAGGCCAAGGCTGGAAGATGTCAGGTCTGGAGCGCGAGCTTGCCTCAATTTGGCG 409  
 DB 224 ArgArgGlyArgAlaGlyArgAlaAlaAlaAlaAlaGlnArgArgLeuGlyArgAlaAla 243  
 QY 410 CCCCCTGAGGGCAGCGGACGCTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 469  
 DB 244 ProLeuAlaAlaGlyAlaAlaArgSerArgGlyAlaGlyGlnProArg----- 259  
 QY 470 TCCACACGTTCATCGACGAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 524  
 DB 260 -----GlyThrArgArgArgArgThr-AspProArgLeuProAlaGlyGlyGly 275  
 QY 525 -----CGTTGCGCACTTCTCTGAGCGGCAAGTACGACCTCTCTCTCTCTCTCTCTCTCTCT 577  
 DB 275 YAspArgAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 292  
 QY 578 CTTGGTGGCGGCTTCT 637  
 DB 292 aLeuArgAla-----LeuLeuArgArgAlaLeuArgTyrArgAlaAlaAlaAlaAlaAla 310  
 QY 638 CCGA-----GCGCGCGCT 673  
 DB 310 nArgGlyAspProArgProProAlaAlaAlaProGlyLeuProArgAlaAlaAlaGlyArgGly 330  
 QY 674 CGGT-----GTGGGAGCAGAACACCC 694  
 DB 330 sArgProValValProLeuProProProValLeuArgProAlaAlaAlaGlyAlaSerLeuAl 350  
 QY 695 AGGTTGATCCCT 754  
 DB 350 aArgLeuArgSerValArgArgLeuProAlaProAlaAlaAlaAlaAlaAlaAlaAlaAla 370  
 QY 755 TGCCCGAGGTTGGCCAGCGCTCTGACGAGGAGCGCTGGAGCGCGCGCCCAT----- 806  
 DB 370 o-----TrpProAlaArgProGlyGlyAlaGlyAlaAlaAlaAlaAlaAlaAlaAlaAla 387  
 QY 807 -----GTTCCCGCCCACTTCTCTCTA-----CACGCACTGCTGGAGGA----- 845  
 DB 387 YArgGlyGlyGluProGlyAlaGluProValGlyArgAlaAlaAlaAlaAlaAlaAlaAlaAla 407  
 QY 846 -----CCCCGAGCGGATATGAGGTGATGGAGATCAACCCCAAGCAGCAGCTGCTGA 898  
 DB 407 sArgHisProAlaAlaLeuGluAspGlyProAlaAlaGlyGlnProProGlyGluHisAlaAla 427

QY 899 CCCTGACTAGCGCGCTGCAATGCCCTGAACCTGCT-----GGTGAGG-----GG 946  
 DB 427 aProAsp-----ProAlaValArgLeuGlyAlaGlyAlaGly 439  
 QY 947 CCGGCCAGGTGGTTCGCTGGAGCTGCAACCCCGCGAGTCGCGCTTCTGAGCTGAAGA 1006  
 DB 439 yLeuProAlaGlyCysArgGlyArgThrGlyArgProThrGlyAlaPheProAlaGlyAla 459  
 QY 1007 AGGTGGCCATTCAGCAGCTGGAGTTTGAGGACGTGTGCGAGCT----- 1049  
 DB 459 aGlyArgValArgAlaAlaAlaArgProAlaGlyProValAlaGlyProGluArgGlyAspArg 479  
 QY 1050 -----GTTCCGCGAGGCGGTGCAACCCCGCATTCAGGAGCTGTACGAGA 1093  
 DB 479 gAlaArgProArgArgTyrArgGlnGlyArgGlyAlaLeuProArgGlyLeuAlaGlyPr 499  
 QY 1094 AGAAGCTGGCGCCCTT-----CCTGTC-----GCAACACGAGCCA----- 1127  
 DB 499 oArgArgGlyAlaLeuTrpHisProProAlaMetProValAspAlaGluGlnProGlyGly 519  
 QY 1128 -----CAACTTCTGTCCCAAGCGCTCTGCTACTTCCAGCAGCGCT-----GTACT 1174  
 DB 519 yAspProTrpArgLeuLeuAlaGlyAlaGlnLeu-----GlnProArgArgProGlyValArg 538  
 QY 1175 ACCAAGCGCGCATGGGCAAGCTGTGCTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1234  
 DB 538 gProAlaLeuArgGlnProAlaVal-----ArgGlyThrGlyPr 551  
 QY 1235 TGGCAGAGCGCTCAAGCGCT-----CGCAGCGCGCCCAATGAGGAGC 1282  
 DB 551 oLeuArgProArgProArgProGlyAlaAlaArgArgGlyAlaAlaArgArgGlyGly 571  
 QY 1283 AGCGCGCTCTGTGGA-----CAGCAACATGCTCATCCACTTGTGTGAAGAAGC 1330  
 DB 571 yAlaProGlyProGlyThrProProAlaProAlaGlyAlaThrProLeuArg-----Gly 589  
 QY 1331 GGCCCAAGCGCTGGTGGTGTGCTCAAGTTCGTGAGCTGCTGCTCTTCAACAGG 1390  
 DB 589 yAlaArgAlaAlaProAlaValProArgLeuProAlaGlnProGlyAla----- 605  
 QY 1391 CCGTGTGTGGTTCGCGCGCGCTGCGGCGAGCAGTACGCGCTGATCAAGCGGAGC 1450  
 DB 606 -----AlaAlaGlyArgSerAlaGlnValAspGlnAlaGlyAs 618  
 QY 1451 GCATCCCATTTAGAACTACATCGCGCGCACCATTGA-----CGGCTGCGGAGGA 1501  
 DB 618 nArgGlyAsnProGlnLeuProGlyCysGlnProGlyAspArgLeuLeuArgProGlyGly 638  
 QY 1502 ACTCGCAGCTGCGCAAGCAGAACTTCTTACTACAACTGCT-----CACCGGCA 1552  
 DB 638 nProGlyArgAlaProGlyGlnLeuArgGlyGlyValArgProAlaTrpArgArgAl 658  
 QY 1553 AGTT-----CCTGCGCAGCAACTGCGCCACCTACCTGCGCGAGGCGG 1594  
 DB 658 aValAspAlaCysLeuGlyTyrProAlaAspLeuLeuPro-----Gly 672  
 QY 1595 CTTTCGCGCCCTCAAGAGTGGCT-----GTGGACAACCTCAGCTCTCCACCA 1645  
 DB 672 yGlyAspHisProAspGlnValArgThrLeuAlaGlyProGlyProAspArgThrGlyArg 692  
 QY 1646 ACTTCTTCATGGAGGAGCTCAAGCGCGCACTACCAAGGTGATTCTGATGACCACG 1705  
 DB 692 gArgLeu-----AlaThrAlaProGlyArgHisLeu-----Ar 703  
 QY 1706 TGAGCTGCTGGATATGCCCGTGGCCCAACGAGCTGGCGAGTGTGCTGCGCAGCAGGTTG 1765  
 DB 703 gTrpAlaGlyGlyHisAlaAlaGlyMetGlnSer-AlaAlaAlaAlaAlaCysGlyVala 723  
 QY 1766 CGCGCGCGCATCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1816  
 DB 723 laProGly-Arg-----ProGlyAlaProArgThrProGlyGlySerArgProAla 740





```
QY 1237 CCAGTCCAGCACCA-----CGGCCAGGCACTGCGACGCCAGC 1199
Db   :|||:|
324 ArgLeuProAlaAlaAspArgHisGlyAlaAlaLeuGlyProGlyThrAlaAlaArgArg 343
QY 1198 ACAGCTTGGCCATCCGCCCTGGT-----AGT 1172
Db   :|||:|
344 ThrHisArgProTrp---ProGlyArgGlyAlaGlyAspAlaLeuLeuArgProAsp 362
QY 1171 ACAGCGCGTGTGAAGTACAGAGC-----GCTTGACC 1136
Db   :|||:|
363 AlaGlyGlnAlaAspArgLeuArgGlyAspProArgArgSerProAlaGlnAlaAspPro 382
QY 1135 AGAGTTGTGGCTGTTGGCAGCAGG-----AGSGCGCCAGCTTCTTCGTCACAGCT 1082
Db   :|||:|
383 ArgAlaGlyAlaLeuArgAlaAlaGlyArgArgGlnProAlaValProArgGlnPro 402
QY 1081 CCTCAATGCGGGGTGCACGC-----CCTCGCCGCAACAGCT 1046
Db   :|||:|
403 ProGlyAlaSerGlyLeuArgArgArgGlyAspHisArgLeuHisArgThrLeu 422
QY 1045 GCCACAGCTCTCAAACTCCAGCTGCTGAATGGCCACCTTCTCAGCTCCAGAAAGCCGC 986
Db   :|||:|
423 ArgArgArgProGln-----ProAlaThrAlaProAlaGly--- 434
QY 985 ACTGGCGGGGTTCAGTCCACCCAGCACCACTCGCGCGCCCTGCACACAGCAGT--- 929
Db   :|||:|
435 ---ArgArgGly-----ThrGlyProGlyArgArgProAlaLeuProGlyArg 449
QY 928 ---TCAGGCGATTGCGCGCGCTAGTACGGGTGAGTGTCTTGGGGTTGATCT 872
Db   :|||:|
450 ArgArgGlyGlnCys----- 454
QY 871 CCATCACTCCATATCCGGCTCGGGTCTCTCCACGACTGCTGTACAGGAAGTGGCG 812
Db   :|||:|
455 ---AlaProAlaArgProGlyArg-----LeuAlaGlnArgArg 467
QY 811 GGACATGGCGCGCTCCAGCGCTCTCGTGCAGGCGTGGCCACACGCTGGCGAGGC 752
Db   :|||:|
468 GlyThrValGluAspArgProGluAlaArgArgThrAlaArgGlyAspArg- 484
QY 751 GGCCAAATCCACAGTAGTAGGGGGCGCGCAGCCAGCGATAGGGGATCGAAACCTGGG 692
Db   :|||:|
485 -----ProGlyArgProArgLysArgSerProAlaAlaThrAlaGlyAla---ProGly 501
QY 691 TGTCTGTCTCCACACGGGCTCCAGCT----- 665
Db   :|||:|
502 AlaThrArgAspArgArgAlaProAlaArgLeuArgArgThrLeuGlyGlnArgArgThr 521
QY 664 -----TCTGCTCCA-----GTTAGCGC-----GGCGCTCGG 638
Db   :|||:|
522 ArgArgHisSerProProArgValProArgGlyArgArgAlaProLeuGlyThrArg 541
QY 637 GGCCGATGTCAATGTTGTCGATGTCGAAGATCGATCGCCAGAGAAACGGCGCAGCAGG 578
Db   :|||:|
542 ProArgArgGlnProArgThrGlyArgGlyHisProArgArgArgGlnArgProGlyArg 561
QY 577 -----GCATCTGGCCCA-----GGGCGAGTCTGT 554
Db   :|||:|
562 ArgGlnPheArgHisAlaGlnGlyAlaAspGlyArgArgHisArgArgGlyAlaGlyArg 581
QY 553 ACTTGGCGCTCAGTAGAAGTCGGCAACGCCACCGCGCTTGGGACAGGTACGAGC 494
Db   :|||:|
582 -----ArgGlyArgThrGlyArgGlnGlyProThrAla---AlaGlyAlaGlyGlyAsn 598
QY 493 AAGCCTGTGTCGATGACGTTGTGGAACGGTGGGAATCATCGTAGCGAGTAGGAGAGGTGA 434
Db   :|||:|
599 GluAspGly----- 601
QY 433 TGAGCGTCCGGTGCCTCAGGGGGGGCAAAATGGCAAGCTTCGGCTCCACGACTGGA 374
Db   :|||:|
602 -----ThrProAlaGluGlyArgSerArgArgGlyProProArgProGly 617
QY 373 CATTTCTCCAGCCCTTGGCTTGGCTTCTTCTTGGCCACCTCGCACAGCGAGTGGCACA 314
```

```
Db 618 Arg-----ProTrpArgAla---GlyAlaGlnProAlaGlyAlaGlyArgGlyGly 633
QY 313 GGTGCGACCGTAGATGGACTTGAACCTTCCGCCAGGTGATGTAATCAGCCA----- 263
Db 634 SerArgArgLeuArgHisProAlaAlaHisProGlnArgGlnArgGlnGluAsnGlyAsp 653
QY 262 -----TCATATCGACATTTCTCCCGAGTGCACCCAGCTCAACCCAGATGAGGT 212
Db 654 AlaHisValThrAlaArgGlnAspPro---ValHisHisArgArgGlnProArg- 670
QY 211 TCGAGCGCTCGGCCAGC----- 194
Db 671 -----HisArgProGlyAspSerProAlaArgArgAlaArgTrpArgGlnProGlyAsp 688
QY 194 ----- 194
Db 689 ArgGlyGluArgArgAlaAlaSerGluAlaGlyArgHisHisLeuGlnArgArgArg 708
QY 193 ---GGCGAGCAAGGCGAGCGCTGGG-----GCCCGTAGAAGC 158
Db 709 ArgGlyArgGlyArgArgArgProGlyProAlaAlaAlaThrArgArgAlaArgAla 728
QY 157 TCTCCAGCGAG-----CAGCGTGTATCGCCCTTCT 125
Db 729 GlyArgGlyGlyHisGlyProArgArgArgThrLeuArgArgArgArgArgAla--- 747
QY 124 TGCTGCGCAACCATATGCGCAGACGGTCAAGTTCATCTTCATGCTGCTGAGCTTGA 65
Db 748 -----GlyGlnGlnArgGlyAspPro----- 755
QY 64 GCTTCTCCAGGAGAAGTTCCTTCTGTGTAGTTCGCGAGCGCGCGCTCACGAC 11
Db 756 -----ProGlyArgGlyGlyGlu---AlaGlyAlaGlnAlaLeuArgProAsp 770
```

Search completed: March 14, 2006, 01:47:19

Job time : 80 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: March 14, 2006, 01:47:31 ; Search time 3 Seconds  
(without alignments)  
3612.968 Million cell updates/sec

Title: US-10-620-914-44  
Perfect score: 3619  
Sequence: 1 atggggctgggtcgtgacgg.....gcgccaagaaggacaactaa 1947

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 161667 seqs, 27834885 residues

Total number of hits satisfying chosen parameters: 323334

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh  
-Q=/abs/ABSSWEB.spool/US10620914/runat\_13032006\_101939\_5686/app\_query.fasta.1  
-DB=Published Applications AA New -QFMT=fastea -SUFFIX=n2p.rapbn -MINMATCH=0.1  
-LOOPC1=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -HOST=abs05h  
-USR=US10620914.OCGN 1.1.8 @runat\_13032006\_101939\_5686 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WAEN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications AA New:

1: /cgn2\_6/ptodata/1/pubpa/US08\_NEW\_PUB.pdb:  
2: /cgn2\_6/ptodata/1/pubpa/US06\_NEW\_PUB.pdb:  
3: /cgn2\_6/ptodata/1/pubpa/US07\_NEW\_PUB.pdb:  
4: /cgn2\_6/ptodata/1/pubpa/PCT\_NEW\_PUB.pdb:  
5: /cgn2\_6/ptodata/1/pubpa/US09\_NEW\_PUB.pdb:  
6: /cgn2\_6/ptodata/1/pubpa/US10\_NEW\_PUB.pdb:  
7: /cgn2\_6/ptodata/1/pubpa/US11\_NEW\_PUB.pdb:  
8: /cgn2\_6/ptodata/1/pubpa/US60\_NEW\_PUB.pdb:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	294.5	8.1	386	7 US-11-096-568A-21828	Sequence 21828, A
2	279.5	7.7	386	7 US-11-096-568A-21828	Sequence 21828, A
3	275	7.6	1886	6 US-10-515-868-8	Sequence 8, Appl
c 4	270	7.5	495	7 US-11-182-016-31	Sequence 31, Appl
5	267.5	7.4	428	7 US-11-096-568A-19503	Sequence 19503, A
6	267.5	7.4	493	7 US-11-096-568A-11657	Sequence 11657, A
7	267.5	7.4	544	7 US-11-096-568A-19501	Sequence 19501, A
8	267	7.4	1076	6 US-10-131-826A-219	Sequence 219, App
9	267	7.4	1076	6 US-10-973-115B-219	Sequence 219, App

10	262.5	7.3	1742	7	US-11-182-016-23	Sequence 23, Appl
11	261	7.2	413	7	US-11-096-568A-20771	Sequence 20771, A
12	262	7.2	1356	6	US-10-894-592-3	Sequence 3, Appl
c 13	258.5	7.1	766	7	US-11-096-568A-19864	Sequence 19864, A
14	257.5	7.1	1717	7	US-11-182-016-20	Sequence 20, Appl
15	256	7.1	1652	7	US-11-241-631-1	Sequence 1, Appl
16	255	7.0	375	7	US-11-096-568A-23618	Sequence 23618, A
c 17	254.5	7.0	459	7	US-11-096-568A-21887	Sequence 21887, A
18	253	7.0	2508	7	US-11-241-631-7	Sequence 7, Appl
19	253	7.0	2544	7	US-11-241-631-3	Sequence 3, Appl
20	253	7.0	2601	7	US-11-241-631-9	Sequence 9, Appl
21	251.5	6.9	1128	7	US-11-241-631-11	Sequence 11, Appl
22	250.5	6.9	758	7	US-11-096-568A-14593	Sequence 14593, A
c 23	249.5	6.9	493	7	US-11-096-568A-11657	Sequence 11657, A
c 24	248.5	6.9	413	7	US-11-096-568A-20771	Sequence 20771, A
c 25	247	6.8	428	7	US-11-096-568A-19503	Sequence 19503, A
c 26	247	6.8	544	7	US-11-096-568A-19501	Sequence 19501, A
27	246.5	6.8	1917	7	US-11-241-631-5	Sequence 5, Appl
c 28	242.5	6.7	373	7	US-11-096-568A-24110	Sequence 24110, A
c 29	240.5	6.6	369	7	US-11-096-568A-24111	Sequence 24111, A
30	239	6.6	558	7	US-11-096-568A-26217	Sequence 26217, A
c 31	238	6.6	449	7	US-11-096-568A-20255	Sequence 20255, A
32	238	6.6	575	6	US-10-980-388-46	Sequence 46, Appl
33	236	6.5	382	7	US-11-096-568A-26984	Sequence 26984, A
34	235.5	6.5	544	6	US-10-980-388-40	Sequence 40, Appl
35	235	6.5	442	7	US-11-096-568A-19628	Sequence 19628, A
36	234.5	6.5	535	7	US-11-096-568A-25889	Sequence 25889, A
c 37	233.5	6.5	512	7	US-11-096-568A-25890	Sequence 25890, A
c 38	232.5	6.4	325	7	US-11-096-568A-24112	Sequence 24112, A
39	232.5	6.4	459	7	US-11-096-568A-21887	Sequence 21887, A
40	232.5	6.4	579	7	US-11-096-568A-24333	Sequence 24333, A
41	232	6.4	369	7	US-11-096-568A-24111	Sequence 24111, A
42	232	6.4	373	7	US-11-096-568A-24110	Sequence 24110, A
c 43	231.5	6.4	375	7	US-11-096-568A-23618	Sequence 23618, A
44	231.5	6.4	526	7	US-11-096-568A-16997	Sequence 16997, A
45	230.5	6.4	702	7	US-11-096-568A-14646	Sequence 14646, A

#### ALIGNMENTS

RESULT 1

US-11-096-568A-21828

; Sequence 21828, Application US/11096568A

; Publication No. US20060048240A1

; GENERAL INFORMATION:

; APPLICANT: Alexandrov, Nikolai et al.

; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides

; FILE REFERENCE: 2750-1592PUS2

; CURRENT APPLICATION NUMBER: US/11/096,568A

; CURRENT FILING DATE: 2005-04-01

; NUMBER OF SEQ ID NOS: 34471

; SEQ ID NO 21828

; LENGTH: 386

; TYPE: PRT

; ORGANISM: Zea mays subsp. mays

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (1)..(386)

; OTHER INFORMATION: Ceres Seq. ID no. 12406424

US-11-096-568A-21828

Alignment Scores:			
Pred. No.:	1.69e-10	Length:	386
Score:	294.50	Matches:	127
Percent Similarity:	37.4%	Conservative:	26
Best Local Similarity:	31.1%	Mismatches:	117
Query Match:	8.1%	Indels:	139
DB:	7	Gaps:	28

US-10-620-914-44 (1-1947) x US-11-096-568A-21828 (1-386)

QY 1684 TGGTAGGTGGCGGCTTTGAGCTCTCCATGAAGAAGTTGGTGG-----AGA 1637

```
Db 35 TrpArgCysAlaAlaThrArgProGlnArgHissargTrpProAspArgArg 54
QY 1636 CGG-----TCAGGTTGTCACACCGCCACTCTTTAGGGTGGGAAGCGCCCTCGCGCA 1583
Db 55 ArgProGlyProGlyArgAlaProArgGlyAlaAlaGlyProArgArgPro----- 71
QY 1582 GGTAGGTGGGCGCATTGTCG----- 1562
Db 72 --GlyGlyGlySerThrArgArgLeuLeuAlaAlaArgGlnProGlnMetProArg 90
QY 1561 ---CCAGGAACCTCCGCTGAGGAGTGTAGTAGAGAGTAGTCTGCTGTCGACGTCGCG 1505
Db 91 GlySerGlyAlaSerArgProGlySerCysSer-----SerArgGly 104
QY 1504 AGTTCTCCGCGCACCGCTGTCATGTCGCGCGATGTAGTCTCAATGGGATCGCTCG 1445
Db 105 SerThrGlyProArgArg---TrpProAla---AlaGlyGlyArgTrp----- 118
QY 1444 CTTGATCAGCGCGTACTGTTGCCGCGCA-----CGCGCGCGCGAACC 1400
Db 119 -----ArgThrProCysProCysAlaSerProSerProArgArgAlaArgPro 134
QY 1399 ACAGCAGCGCTTGTGAAGAGCACCGGCTCAGCACTTGCAGCAAGCCACACACCGAGC 1340
Db 135 ThrAlaAlaGlyAlaThrSerThrGlyCys-----SerThrAlaThrAlaAla 151
QY 1339 ---GCTTGGGCGCGTCTTTCAGAGTGTGATGAGCATGTTGCTGTCCACAGCGCGCT 1283
Db 152 ArgThrTrpProArgArgAlaGlyThrGlyCysThrThrCysSerPro----- 167
QY 1282 GCTCTCCATTGTCGCGCGTGTGGGAGCGCTGACGCTTGCCTCCAGTCCAGCACCA 1223
Db 168 -----ArgSerSerProLeuProSerPro 175
QY 1222 -----CGSCCAGGC----- 1190
Db 176 ProThrAlaProArgArgProGlyProProProArgTrpAsnAlaAlaSerArgGlyTrpThr 195
QY 1189 CCA-----TGCGCGCTGTAGTACAGCGCGTGTGGAAGTACAGCA----- 1148
Db 196 ProArgArgCysArgProAlaGlyAlaProProAlaAlaAlaThrArgThrSerAlaThr 215
QY 1147 -----GGCGCTGGACAGAGTGTGGCTGTGCTGTCGCGCA 1112
Db 216 ThrTrpAlaProProTrpTrpProSerTrpArgSerAlaAlaSerTrpTrpProThr 235
QY 1111 GGAAGCGCGCAGCTTCTCTGTCAGCTCTCAATGC-----GCGGTGCGCGCCT 1058
Db 236 -----ValAlaThrAlaAlaArgCysAlaAlaGlyAlaAlaArg 249
QY 1057 GCGCGAAGCTGCCACAGCTCTCAACTCCAGCTGCT-----GAATGG----- 1013
Db 250 ArgArgCysArgCysArgArgThrThrSerProThrAlaProThrSerTrpSerGlySer 269
QY 1012 ---CCACCTCTTTCAGTCCAGAGCGCGACTGCGCGGGTGTGAGTCCACCGACCA 956
Db 270 ArgProArgAlaAlaAlaSerSerGlyArgAlaAlaArgAlaCysTrpAlaSerSerPro 289
QY 955 CTGGCGCGCGCTGCCACAGCAGTTCAGGCGATTGCGCGCGCTAGTCAGGGTCA 896
Db 290 -----CysProAlaProSerGlyThrAlaThrSerSerArgThr-----Ser 303
QY 895 GCACCGTGTCTTGGGGTTGATCTCCATCACTCCATATCCG-----GCTCGGGTCTCT 842
Db 304 ArgProCysProArgSerArgSerProThrAlaProThrProThrSerAlaSerSerSer 323
QY 841 CCCACGACTCGGTGTACAGAGAGTGGCGGGAACATGGCGCGCGCTCCAGCGCTCTCT 782
Db 324 ProAlaThrAla-----SerGlyThrTrpLeuAlaThr-----ArgPro 336
QY 781 CTGTCAGGCGCTGGCCCAAGCTGGGCGAGCGCGCAATCCACAGCTAGTAGGGGCGCGCA 722
```

```
Db 337 ArgAla-----ArgTrp-----Arg 341
QY 721 GCCACGCGCAGTAGGCGATCGAACCTGGG-----TGTTCTGCT 683
Db 342 AlaProAlaSerAlaGlyGlyProGlyGlyAlaProArgProProArgCysSerPro 361
QY 682 CCCACACGCGCTCCAGCTTCTGCTCCA 656
Db 362 SerTrpProSerProGlyAlaAlaPro 370

RESULT 2
US-11-096-568A-21828
; Sequence 21828, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE OF INVENTION: Therby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 21828
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURES:
; NAME/KEY: misc_feature
; LOCATION: (1)..(386)
; OTHER INFORMATION: Ceres Seq. ID no. 12406424
US-11-096-568A-21828

Alignment Scores:
Pred. No.: 1.32e-09 Length: 386
Score: 279.50 Matches: 126
Percent Similarity: 36.4% Conservative: 31
Best Local Similarity: 29.2% Mismatches: 123
Query Match: 7.7% Indels: 151
DB: Gaps: 22

US-10-620-914-44 (1-1947) x US-11-096-568A-21828 (1-386)
QY 711 CGTGGCTGGCTGGCGCCCTACTACGTGTGATGGCGCGCTGCCAGGTTGGCA 770
Db 37 ArgCysAlaAlaThrArgProGlnArgHissargTrpProAspArgArgArgPro 56
QY 771 CGCCCTGCACGAGAGCGCTGGAGCGCCGCCCATGTTCCGCGCCACCTTCTCTACAC 830
Db 57 GlyProGlyArgAlaProArgGlyAlaAla----- 66
QY 831 GCAGTGTGGGAGGACCCCGAGCGGATATGAGGTGATGAGATCAACCCACAGACAC 890
Db 67 -----GlyProArgProGlyGlySerThr----- 76
QY 891 GGTGTGACCTGACTAGCGCGCGCTG-----CAATGCCCTGAA 929
Db 77 -----ArgArgLeuLeuAlaAlaArgGlnProGlnMetProArg 90
QY 930 CTTGTGTGACGGGGCGCGCGAGTGTGCTGCTGCTGCAACCCCGCGCAGTCCGC 989
Db 91 GlySerGlyAla---SerArgProGlySerCysSer-----SerArgGly 104
QY 990 GCTTCTGAGCTGAAGAAGTGGCCATTCAGCAGCTGGAGTTTGAGGAGCTGTGGCAGCT 1049
Db 105 SerThrGlyPro---ArgArgTrpProAlaAlaGlyArgTrpArgThrProCysProC 124
QY 1050 GTTCGGCGAGGCGTGCACCGCGCATTCAGGAGCTGTACGAGAGAAGCTGGCGCCTT 1109
Db 124 ysAlaSer-----ProS 128
QY 1110 CTGTGCGGAACACCGACCACTTCTGTGTCGAAGCGCTCTGTACTTCCAGCAGCGCT 1169
Db 128 erProArgAlaArgProThr-----AlaAlaGlyAlaThrSerThrGlyC 145
```









1139	QY	GACCAGAGTTGGCTGG---TTTGGCAGAGNAGGCGCGAGCTTCTTCTCGTACAGC	1083
228	DB	 pThrProSerProSerAlaLysSerProSerGlyAlaArgGlnTh	244
1082	QY	TCCT-----	1068
244	DB	 rProAlaHisPheGlySerLeuLysProLeuLeuAlaAlaProThrSerCysSerG	264
1067	QY	TGCACGCCCTCCGGAAACAGCTGCCACACCTCTCAAACCTCCAGCTGCTGAATGGCCACC	1008
264	DB	 yGlyTyrProArgHisProAla-----ProAspSerAlaAlaProGlyValGlnPr	281
1007	QY	T-----TCCTTCAGCTCCAGAAGCGCGAGCTGCGCGGGTTCAGTCCACCGACACACC	954
281	DB	 oHisAlaAlaHisAlaProGlyAlaProAlaProAlaAlaValProProArgProAl	301
953	QY	TGGCGGCGCCCTCGCA-----	938
301	DB	 aAlaPheProProAlaAspGlySerProProSerLeuLeuTyrProArgSerCysLe	321
937	QY	-----	927
321	DB	uTyrGlnProPheSerGlnThrArgHisTyrSerSerGlyThrGlnSerProLeuGlyPr	341
926	QY	AGGCGATTGCAGCGCGCTAGTCAGGGTCAGACACCGTGT-----	887
341	DB	 oGly-----ValProArgProGlySerGlyHisSerProCysGluSerCysSerTrpHisLe	360
886	QY	-----CCTTGGGGTTGATCT-----CCATCACCTCCATATCCGGCTCGG--	848
360	DB	 uLysProTyrProProSerProCysThrGlnAlaProHisProProArgProValArgTr	380
847	QY	-----GGTCTCCACAGACTCGGTGACAGNAGGTGGCGGGAACTATGGCGCGCGCG	795
380	DB	 pSerHisGlyProProSerGlySerTyrPro-----TrpCysArgGlyTyrHisArgLe	398
794	QY	TCCACGCGCT-----CCTCGTCAGGGCGTGGCCACAGCTGGCGGACCGCGGCCA	747
398	DB	 uProSerAlaHisArgSerArgProArgLeuSerSerGlyGlnIleTyrAla-----	415
746	QY	ATCCACACGTAGTAGGGGGCGCGACGCCACGGCACCGTAGGGGATCGAAACCTGGGTTC	687
416	DB	---ValGlnSerTyrGly---ProSerLeuCysArgArgTyrThrSerProSerArgCy	433
686	QY	TGCTCCCAACACGGCTCCAGCTTCTGCTCCAGTAGCGCGCGCTCGGGGCCGATGTCA	627
433	DB	 sAlaProProProSerProGlyHisProProLeuCysGlnProArgGly---CysHi	452
626	QY	ATGTTGTCGATGTCGAAGATCGATCGCCAGAAAGACGGCGCACCGAGCATCTGGCGC	567
452	DB	 sCysCys-----	454
566	QY	AGGGCGAGTCTGPACTTGGCGCTCAGCTAGAAAGTCGGCAACGCCCAACGAGCGCTTGG	507
455	DB	 -----CysLeuHisArgArgGluProProSerArg-----	463
506	QY	GACAGTAGCAGCAACCTGGTTCGATGACGTTTGTGAAACGGTGGATCATCTGACCGAG	447
464	DB	 -SerGlyThrSerArgPro-----	469
446	QY	TAGGAGAAAGTGTAGCGTTCGGGTGCCCTCAGGGGGCGCAAAATTGGCAAGCGTCGGCC	387
470	DB	-----	474
386	QY	T-----CCACGACTGGACATTTCTCCAGCCCTTGGCCCTTCGCCT	347
474	DB	 oLeuAlaAlaLeuAlaArgSerGlySerProProTyrProAlaPro	491

## RESULT 5

US-11-096-568A-19503

US-11-096-568A-19503  
: Sequence 19503. Application US/11096568A

; sequence 19503, Application US7/

; PUBLICATION NO: US20  
; GENERAL INFORMATION:

```

; APPLICANT: Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; TITLE OF INVENTION: Therby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 19503
; LENGTH: 428
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(428)
; OTHER INFORMATION: Ceres Seq. ID no. 12373378
US-11-096-568A-19503

```

**Alignment Scores:**

Pred. No.:	6,878-09	Length:	428
Score:	267.50	Matches:	126
Percent Similarity:	41.6%	Conservative:	30
Best Local Similarity:	33.6%	Mismatches:	103
Query Match:	7.4%	Indels:	116
DB:	7	Gaps:	30

US-10-620-914-44 (1-1947) x US-11-096-568A-19503 (1-428)

205	CGCTCGAACCTCATCTGGGTGTGACTGGGTGGT---	GGCACTCGGGGAGAAATGTCGATATG	261
131	ArgGlyAsnIleLeuSerTyrAsnLeuGlyGlyHisGly-LeuGlnArg-		146
262	ATGGCTGATTACATCATGACCTGGCGAAGTTCAAAGTCCATCTACGTGG---	TCGACCTGTGC	318
147	-----ArgAlaSerSerProSerIleTrpProAlaThrCysSe		159
319	CACCTCGCTGTGCAGGTGCCAAGAAGAGCGCAAGCGCTGGAAGAATGTCAC		378
159	r-----ArgProAlaAla-		164
379	GTTCGTGGAGCGCAGCGCTTGCCCAATTTGGCCCGCTGAGGGCAACCGCGAGCTCATCACC		438
164	galaTrp---ProTrpTrpSerAlaAlaArgPro-----SerProSerArgGlyThrPr		181
439	T-----TCTCCTACTCGCTCAGCATGATTCACCGTTCCACACAGCTCATCGACCAG		489
181	oGlySerGlyAlaProCysSerSerArg-----ThrProSerSerGlyPr		196
490	GCTTGCTCGTACCTGT-----CCCAAGACGGC		516
196	oAlaAlaArgArgCysCysProThrGlyAlaAlaThrSerThrAlaProSerThrSe		216
517	CTGG-----TGGCGCTTCCGAGCTTCTAGTGAGCGGCAAGTACGACCTGCCCTCGCC		570
216	rTrpSerThrTrpCysAlaAlaArgThr-----ArgAlaAlaArgThrThrAlaProSerAl		233
571	CAGATGSCCTGTGTGGCGCGTTTCTTCTGCGATCGATCTTCGACATCGACACATTGAC		630
233	aProCysThrArgArgArgThrSerSerGlySerArg-----		245
631	ATTCGCCCGCGAGCGCGCCCTACTCTGGAGCAGAAAGCTGGAGCGCGTGTGGGAGCAGAAC		690
246	-AlaCysProSerAlaAlaThrCysTrp-----ArgTrpAlaAlaThrArgSerArgPr		263
691	ACCCAGGTTTCA--TCCCTCTACTGTCGCGTGTGGCGCGCCCTACTACTAGTGTGGATT		747
263	oThrSerProArgTrpGlyProTrpCys-----CysProSerProSerSerCysSerSe		281
748	GGCC---GCCTCCCCAGCGTTGGCCACGCGCTGACGAGGAGCGGTGGAGCGCGCGCC		804
281	rSerArgAlaCys-----Cys-----SerAlaThrCysSerArgPr		293
805	ATGTTTCCCGCCCACTTCTGTATCACGCAAGTCGTGGGAGGACCCCGAGCGCGATATGGAG		864











QY	1364	TCGTGAGCTGGTCTTCAACAGCGCGTGTGGTTCGGCGCGGCGTCCGGCA	1423	QY	8	CGGTCGTGACGGCGCTCGGAGTACACCAAGAAAGAACTTCTCCCTGGAGAAGCTCA	67
Db	486	AlaAlaAlaGlyCysCysAlaThr-----CysAlaGlyThrGlyAlaAlaCysThrGly	503	Db	439	ArgThrAlaSerThrAlaSerArgSerLeuPro---ArgAlaSerCysSerArgSerSer	457
QY	1424	AGCAGTACGGCTGATCAAGCGGAGCGGATCCCATTTGAGAACTACATCGCGCGCACCA	1483	QY	68	AGCTACGACGATGAAG-----	85
Db	504	GlyAla-----ThrAlaThrThrCysThrThrCys	517	Db	458	CysThrAlaAlaProAlaAlaArgHisSerGluLeuSerSerGlyProAlaSerLeuArg	477
QY	1484	TGACGCGCGTGGCGGAACTCGACGTGGCGAAGCACTACTTCTACTACA-----	1537	QY	86	-----ATGACCTGACGCTCTCGCCATATGTTGGTCGGCAGCAAGAGCGCATG	136
Db	518	CysThrGlyAlaThrThrAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla	537	Db	478	SerSerSerArgSerArgProTyArgValSerCysSerArgGlyCysCysArgAla---	496
QY	1538	-----ACTGCTCACCAGGAGTCTCGCGGACAACTGCGCCACCTACCTCGCGGAGCGG	1594	QY	137	ATCACGCTGCTCGCTCGAGAGCTTCTACGGGCCCCAGGCGCTG---CCTTTCTCTCCC	193
Db	538	AlaAlaAlaGlyCysAlaGlyThrCysAlaGlyThrAlaAlaAlaCysCysAlaAlaAla	557	Db	497	-----ArgGlyArgGlyProAlaSerMetGluProLeuLeuLys	509
QY	1595	CCITCGCCACCCTCAAGAGTGGCGTGGTGGACAACTGACCGTCTCCACCAACTTCTTCA	1654	QY	194	GCCTGGCGGAGCGCTCGAACCTCATCTGGGTTGACCTGGGTGGTGACCTGGGAGATG	253
Db	558	GlyCysCysAlaAlaGlyThrAlaCysAlaThrThrGlyAlaThrThrThrAla---	576	Db	510	GlyLeuSerAlaAlaArgProPheSerSerLeu-----	520
QY	1655	TGAGGAGCTCAAGCGGCACCTACACCAAGTGATTCTGATGACGACGCTGACTGGC	1714	QY	254	TCGATATGATGGTGTATATCATGACCTGGCGAAGTTCAAGTCCATCTACGTGCTCGACC	313
Db	576	-----	576	Db	521	-----CysArg	522
QY	1715	TGGATATGCGCGTGGCCAGCAGCTGGCGGAGTGGCTGGCCCAAGCAGGTTGCGCGGCGG	1774	QY	314	TGTCCTCACTCGCTGCGAGG-----TGCCA-----AGAAGAAGCGCAAGCCCAAGG	361
Db	577	-----CysAlaGlyThrThrAlaThrThrThrThrThrGlyAlaAlaAlaAlaAla	594	Db	523	CysPheThrTyCysAlaAlaArgLeuAsnTrpSerProTrpArgArgLysGlnAlaPro---	541
QY	1775	GCATCGTCACTCGGCGCT-----CGCCTCCCTCAGCGCGC	1810	QY	362	GCTGGAAGATGTCACGAGTCTGTG-----	388
Db	595	AlaThrAlaAlaGlyAlaAlaCysThrGlyCysThrAlaGlyAlaAlaAlaAlaAlaThr	614	Db	542	-----MetArgArgGlyTrpValAlaGlnProSerSerArgLysCysValArg	558
QY	1811	CCTACGCGAGCTGATCCAGAGCGGCGTTCGACGTGGTGGTGGTGGTGGTGGTGGTGG	1870	QY	389	CCGACGCTGCGCAATTG---CGCCCTGAGGCA-----CCGCGACGCTCA	433
Db	615	GlyThrThrThrAlaThrAlaAlaCysAlaGlyThrCysThrAlaThrThrThrCysThr	634	Db	559	TrpLysLeuArgSerLeuSerArgAlaAlaAlaAlaAlaGlnValProArgSerSer	578
QY	1871	AGGGCTACATGACCGCTGACATGATGATGATGATGATGATGATGATGATGATGATGAT	1912	QY	434	TCACCTTCTCTACTCGCTCAGATGATTCACCGTTCACCAACGCTCATCGACCAAGGTT	493
Db	635	ThrThrThrAlaAlaAlaAlaAlaCysThrThrThrThrThrThrThrThrThrThr	648	Db	579	GlyValSerSerGlyProGly-----GlyGlyThrGlySerLeuLysVal	593
QY	1912	-----	1912	QY	494	GCTGCTACTCTGCCAAG-----ACGGCTGCTGGCGCTG-----	529
Db	648	-----	648	Db	594	AlaThrSerCysSerGlnAlaSerSerSerSerSerSerSerSerSerSerSerSerSer	613
QY	1912	-----	1912	QY	530	-----CCGACTTCTAGTGGCGGCAAGTACGACGCTGCGCC	565
Db	648	-----	648	Db	614	ArgLeuArgSerIleArgGlyAlaProThr-----ProGlnGlyProSerGlyPro	630
QY	1912	-----	1912	QY	566	TCGGCAGATGCGCTGGTGGCGCGCTTCTCTGCGATCGA-----TCTTCGACATCG	619
Db	648	-----	648	Db	631	LeuAlaArg---ProSerSerThrCysSerArgGlyGluHisValLeuSerThrThrSer	649
QY	1912	-----	1912	QY	620	ACAACATTGACATCGGCCCGC-----	640
Db	648	-----	648	Db	650	ValLeuValThrSerArgProArgProArgProArgProArgProArgProArgProArg	669
QY	1912	-----	1912	QY	641	-----AGCGCGCGCTTACCTGGAGCAGAAC-----TGAGC	673
Db	648	-----	648	Db	670	GlyAlaSerSerGlyProSerSerSerSerSerSerSerSerSerSerSerSerSerSer	689
QY	1912	-----	1912	QY	674	CGGTGGGAGAGAACACCCAGGTTTCGATCCCTACGTCGCTGCGCTGCTGCTGCTGCTG	724
Db	648	-----	648	Db	690	GlyAlaGly-----ArgGlyProGlyLeuLeuGlySerSerLysGlu	703
QY	1912	-----	1912	QY	725	-----GCGCCCTTACTACTGCT-----	742
Db	648	-----	648	Db	704	LeuArgLeuGlyGlyLeuArgGlyProHisTrpAlaProProLeuLeuAlaProSer	723
QY	1912	-----	1912	QY	743	-----GGATGGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCG	781
Db	648	-----	648	Db	724	AlaArgArgArgProArgGlyLeuAlaSerSerProArgCysCysTrpIleLeuLeu	743
QY	1912	-----	1912	QY	782	AGGAGCGGTGGAGCG	826



```

Db      1058 SerSerMetArgIleSerSer-----ArgThrArgGlyGly----- 106
Qy      1721 TGCCCGTGGCCACAGAGTGGCCGAGTGCCTGGCCAGCAGGAGTGGCCGGCGGCATCG 178
Db      1070 -----SerProGluSerSerTrpLeuysAlaGlyMetSerAlaGlyGly 108
Qy      1781 TCATCTGGCGCTCGCCTCCCTCAGCCCGCCTACGCCGAGCTGATCCAGAAGCGGGCT 184
Db      1085 SerGlyGlyAlaArgSerSerArgThrAspSerProSerValCysArgSerGlyLeu 110
Qy      1841 TCGACG 1846
Db      1105 SerSer 1106

RESULT 11
US-11-096-568A-20771
; Sequence 20771, Application US/11096568A
; Publication No. US20060049240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Pol.
; FILE OF INVENTION: Therby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 20771
; LENGTH: 413
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(413)
; OTHER INFORMATION: Ceres Seq. ID no. 12387083
US-11-096-568A-20771

Alignment Scores:
Pred. No.: 1,46e-08 Length: 413
Score: 262.00 Matches: 130
Percent Similarity: 38.7% Conservative: 42
Best Local Similarity: 29.3% Mismatches: 138
Query Match: 7.2% Indels: 134
DB: 7 Gaps: 26

US-10-620-914-44 (1-1947) x US-11-096-568A-20771 (1-413)
Qy      284 CGAAGTTCAGTCCATCTACGTGGTGCACCTGGCCACTCGCTGTGGAGTGG----- 337
Db      72 ArgGlnAlaAlaAsnGlnProTrpSerAlaSerArgArgSerCysSerTrpAlaSer 91
Qy      338 -----CCAAGAAGAAAGCGCAAGGCGCAAGGCGTGGAAAGATGTCCAGG 379
Db      92 SerAlaSerThrAlaSerProArgArgSerSerGluAlaArgThrAlaArgSerSerArg 111
Qy      380 TCGTGGAGGCGCAGCTTGCCAAATTTCGCCCCCTGAGGGCACCGCGA-----CGC 430
Db      112 AlaSerArgPro-----ProSerArgCysProSerProArgPheSer 125
Qy      431 TCATCACTTCTCTACTCGCTCAGCATGATTCACCGTTCACACAGTCATCGACACGAG 490
Db      126 SerCysThrSerProSerArgThr-----ArgSerSerProThrSerThr--- 140
Qy      491 CTTGCTCGTACCTGTCCCAAGACGGCCTGGTGGGCGTTGCCGACTTCTACGTGAGCGCA 550
Db      141 -----ProThrThrArgProTrpThrHisProProArgAlaProThrArg 155
Qy      551 AGTACGACCTGCGCCCTGCGCGCAGATGCCCTCGTGTGCGCCGCTTCTTCTGGCGATCGATCT 610
Db      156 SerAlaAlaCys-----CysPro-----Ala 162
Qy      611 TCAGACATCGACACATTCGATCGGCGCCCGAGCGCGCGCTACTGTGAGCAGAAGCTGG 670
Db      163 SerProProThrGlySerProCysSerSerSerSerProThrSerSerArgSerSer 182

```

FILE REFERENCE: 68175			CURRENT APPLICATION NUMBER: US/10/894,592		
CURRENT FILING DATE: 2004-07-20			NUMBER OF SEQ ID NOS: 3		
SOFTWARE: PatentIn version 3.2			SEQ ID NO 3		
LENGTH: 1356			TYPE: PRT		
ORGANISM: Homo sapiens			US-10-894-592-3		
Alignment Scores:			Pred. No.:		
Score: 261.00			Length: 1356		
Percent Similarity: 33.6%			Matches: 133		
Best Local Similarity: 24.3%			Conservative: 51		
Query Match: 7.2%			Mismatch: 183		
DB:			Indels: 181		
			Gaps: 25		
US-10-620-914-44 (1-1947) x US-10-894-592-3 (1-1356)					
Qy	473	ACAAAGTCATCGACGAGGCTTCTGCTACCTGCTCCCAAGAGCGGCTGGTGGCGGTGGCG	532		
Db	54	ThrAlaAlaCysThrThrCysAlaThrThrCys	64		
Qy	533	ACTTCTACGTGAGCGGCAAGTACGACCTGCGCCCTGCGCCAGATGCCCTGGTGGCGCGTT	592		
Db	65	---ThrThrCysAlaGlyThrAlaCysAlaGlyAla---CysAlaGlyThrGlyThr	82		
Qy	593	TCTTCTCGGATCGATCTTCGACATCG	640		
Db	83	ThrThrGlyThrGlyThrGlyThrCysCysThrGlyThrGlyThrGlyCys	102		
Qy	641	AGCCCGCGCTACTCGGAGCAGAGCTGGAGCGG	697		
Db	103	ThrAlaAlaCysThrThrThrAlaAlaGlyAlaGlyCysGlyGly	120		
Qy	698	GTTCGATCCCTACGTGCGGCTGGCGCCCTCTACTAG	739		
Db	121	ThrThrThrCysThrCysAlaGlyAlaAla	138		
Qy	740	---TGTGATGGCGCTGCGCCAGCGTGGCGCCCTGCGCCAGCGCGCTGGCGCGTGG	793		
Db	139	AlaCysCysAlaThrGlyGlyCysAlaCysAlaAlaAlaCysThrGlyThrGlyAla	158		
Qy	794	AGCGCGCGCCATGTTCCCGCCACCTTCTGTACAGCAGCTGCTGGGAGGAGCCCGAGC	853		
Db	159	CysGly	171		
Qy	854	CGGATATGGAGGTGATGGAGATCAACCCCAAGGACACGCTGCTGACCTGACTAGCGCG	913		
Db	172		173		
Qy	914	GCTGCAATCCCTGCAACCTGCTGTGTGAGGGGCGCGCAGGTGGTGTGGTGGACTGCA	973		
Db	174	ThrAla	189		
Qy	974	ACCCCGCGAGTGGCGCTTCTGGAGCTGAAGAGGTGCCATTTCAGCAGCTGGAGTTG	1033		
Db	189		189		
Qy	1034	AGGAGCTGT	1054		
Db	190	---ThrCysCysThrAlaThrThrGlyCysAlaCysAlaGlyAlaAlaAlaThrCysAla	208		
Qy	1055	GGGAGCGGTGCACCGCGCATTTGAGGAGCTGTATAGAGAGAGCTGGCGCCCTTCTCT	1114		
Db	209	GlyAlaGlyCysCysThrCysAlaThrThrCys	220		
Qy	1115	CGCAACCGAGCCACAACTTCTGTCTCAAGCGCTCTGGTACTTCCAGACAG	1165		
Db	221	CysThrThrAlaGlyThrAlaGlyThrGlyAlaAlaGlyCysAlaThrThrGlyAlaThr	240		

Qy	671	AGCGCGTGTGGAGCAGAACACCCAGGGTTCGATCCCTACCTGCTGCGCGCGCC	730		
Db	183	CysSerArgSerProSerProCysSerPro	197		
Qy	731	CCTACTACGTGTGATGGCGCGCTGCGCAGCGTGGCCAGCGCTGCAGAGAGCGCG	790		
Db	198	ProSerThr	207		
Qy	791	TGGAGCGCGCCCATGTTCCCGCCACCTTCTGTATACGCACTGCTGGAGAGCCCG	850		
Db	208	---ProSerArgAlaCysSer	224		
Qy	851	AGCCGGATATGGAGGTGATGAGATCAACCCAGGACAGGTGCTGCTGACTAGCG	910		
Db	225	Pro	234		
Qy	911	CGGCTGCAATGCTGACCTGACCTGCTGTGTCAGGGGCGCGCGTGTGCTGGTACT	970		
Db	235	SerProThrThrProSerProCysSerSerSerSerAlaSerSerSerProArgThrThr	254		
Qy	971	GCAACCCCGCAGTGGCGCTTCTGGAGCTGAAAGAGGTGGCCATTCAGCAGCTGGAGT	1030		
Db	255	Ala	268		
Qy	1031	TTGAGGAGCTGTGGCAGCTGTTGCGCGAGGCG	1063		
Db	269	---ProThrCysSerAlaSerSerThrSerAlaSerSerGlyThrSerProAlaSerSer	287		
Qy	1064	---TGCACCCCGCATTTGAGGAGCTGTACGAGAGAGCTGGCGCTTCTGTCGCAAA	1120		
Db	288	ProCysSerSerArgThrThrAlaAla	301		
Qy	1121	CCAGCCCAACTTCTGTGCTCAAGCGCTCTGCT	1174		
Db	302	AlaArgThrSerSerGlySerSerSerProArgValLeuSerSerSerArgSerThr	321		
Qy	1175	ACGAGCGCGCATGGGCAAGCTGTGCTGGTGTGCTGCACTGCTGGCGCTGGTGGGAC	1234		
Db	322	SerSerSerProSerSerSerAla	330		
Qy	1235	TGGGCAAGCGCTCAAGCGCTCGCCCAAGCGCCCAATGAGGAGCAGCGCGCTGT	1294		
Db	331	---SerAlaProGlySerSerGlyAspProAlaProArgArgGlnGlySer	347		
Qy	1295	GGGAGCAGCAATGCTCATCTTCTGTAAGAACGGGCGCCCAAGCGCTGTGTGGTGT	1354		
Db	348	SerAlaSerSerCysSerProArgSerAlaAlaSerSerCysSerArg	363		
Qy	1355	TGCTCAAGTTCGTGAGCTGTGCTTCAACAGCGCGTGTGTGCTGGCGCGCGG	1414		
Db	364	---SerSerArgArgPro	373		
Qy	1415	TCCCGGCAAGCAGTACGCGCTGATCAAGGCGGAGCGCATCCCATTTGAGAACTACATCG	1474		
Db	374	AlaArgAlaThrThrArgAlaSerThrArgProThrSerPro	391		
Qy	1475	CGCGACCATGAGCGCGTGGCGGAGAACTCGCAGTGGCGGCAAGCAACTACTTCTACT	1534		
Db	392	ArgSer	400		
Qy	1535	ACAAGTGCCTCA	1546		
Db	401	LeuIysAlaSer	404		

RESULT 12

US-10-894-592-3

Sequence 3, Application US/10894592

Publication No. US20060019335A1

GENERAL INFORMATION:

APPLICANT: Du, Chunying

APPLICANT: Yang, Qi-Hang

TITLE OF INVENTION: Compositions and Methods to Promote c-IAP Autodegradation

QY 1166 -----GCTGTACTACAGGGCGGCA 1186  
Db 241 GlyAlaGlyGlyAlaGlyCysAlaGlyThrGlyThrCysThrThrGlyGly 260  
QY 1187 TGGCGAAGCTGTCTGGG-----TGCTGAGTGCCTGGCGGTGGC 1228  
Db 261 ThrAlaAlaCysAlaGlyAlaThrAlaGlyCysAlaCysThrCysThrAlaCysCys 280  
QY 1229 TGGGACTGGGCAAGACCGTCAAGCGCTGCGCAAGCGCCCAACATGAGGAGGAGCGGCC 1288  
Db 281 -----ThrThrThrCysThrCysThrCysThrCysAlaGlyAlaCysCys 295  
QY 1289 GTCTGTGGGAGCAACA---TGCTCATCCACTTCTGTAGNACGGGCCCAAGCGCTGG 1345  
Db 296 AlaCysAlaThrAlaThrGlyCysGlyThrGlyAlaThrGlyAlaAla----- 313  
QY 1346 TGTGCTGTCTGCTCAAGTCTGTGAGCGCTGCTCTTCAACAGCGCGTGTGTGTG 1405  
Db 314 ---GlyCysThrAlaThrThrAlaCysThrGlyAlaAlaThrAlaThrAlaCysThrAla 332  
QY 1406 GCGCGCGGTGCGGGCAAGTAGTACGCGCTGATCAAGCGGAGCGCATCCCATTTGAGA 1465  
Db 333 AlaGlyGlyCysThrGlyThrThrAlaThrAlaCysCysThrThrAlaAlaCysThr 352  
QY 1466 ACT-----ACATCGCGCGCACCATGGAGCGGTGGCGGAGNACTCGCAG---TGC 1513  
Db 353 ThrCysThrCysThrThrAlaCysGlyAlaCysGlyAlaCysAlaAlaThrAlaCys 372  
QY 1514 GCAAGCAGAACTTACTTACTACTGCTCACCAGGAGTCTCTGCGGAGCAACTGCC 1573  
Db 373 AlaAlaGlyThrThrAlaCysThrThrGlyGlyAlaAlaAlaAlaThrGlyAla 392  
QY 1574 CCACCTACTCGGAGGCGGCTTCCGCCCTCAAGAGTGGCGTGGAGCAACTGA 1633  
Db 393 AlaThrThrCysAla-GlyAla----- 399  
QY 1634 CCGTCTCACCACTTCTTCTGAGGAGCTCAAGCGCGCACCTACACCAAGGTATC 1693  
Db 400 -----GlyGlyAla----- 405  
QY 1694 TGATGGACCACTGTGCTGTGATATGCCGCTGGCCCAAGAGTGC----- 1742  
Db 405 aAlaGlyAlaThrGlyAlaAlaGlyThrGlyGlyCysAlaGlyGlyThrGlyAl 425  
QY 1743 -----CGAGTGCCTGCCAAGCAGGT---TGCGCCGGGGCGCATCGTCTGCGGCT 1792  
Db 425 aThrCysAlaThrAlaGlyGlyAlaGlyCysCysAlaGly--AlaGlyCysThrGlyAla 444  
QY 1793 CCGCTCCTCAGCGCGCTTCCCGAGCTGATCCAGAGCGGGCTTCAGCGTGCCT 1852  
Db 445 GlyAlaThrGlyAla-----CysThrThrCysAla 454  
QY 1853 GCATCCGCGCGCGCACTCAGGCTTACATGACCGGTCAACATGTACAGCTCTTCTACA 1912  
Db 455 AlaAlaAlaCysAlaCysCysAlaAlaGlyAlaGlyThrAlaCysThrThrGlyAlaAla 474  
QY 1913 TGGCCCGCGCGAGGGCG 1930  
Db 475 GlyCysThrGlyGlyAla 480

## RESULT 13

US-11-096-568A-19864  
; Sequence 19864, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096.568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 19864

LENGTH: 766  
TYPE: PRT  
ORGANISM: Zea mays subsp. mays  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)-(766)  
OTHER INFORMATION: Ceres Seq. ID no. 12375135  
US-11-096-568A-19864  
Alignment Scores:  
Pred. No.: 2,5e-08 Length: 766  
Score: 258.50 Matches: 192  
Percent Similarity: 32.5% Conservative: 38  
Best Local Similarity: 27.2% Mismatches: 234  
Query Match: 7.1% Indels: 243  
Gaps: 42  
US-10-620-914-44 (1-1947) x US-11-096-568A-19864 (1-766)  
QY 57 GGAGAGCTCAAGCTCAGCAGCATGAAGATGACTGACCGTTCTGCCCATATGTGTT 116  
Db 191 GlyAlaGlyAlaHisGlnHisPro-----ProGlyLeuLeuArgHisProLeu 208  
QY 117 CGCAGCAGAGGCGCGATGATCA-----CGCTCTCGCTGGAGAGTTCTA 164  
Db 209 ArgAspProArgGlyHisHisGluAlaGlnHisArgCysGlnPro-----LeuPro 226  
QY 165 GCGGCGCCCA-----GGCGCTGCGCTTGTGTCGCCCTGGCGGAGCG 206  
Db 227 AlaAlaProGlyHisHisArgValGlyArg-----ProGlyProAla 241  
QY 207 CTC-----GAACCTCATCTCGGTGACCTCGGTGGTGGCACTGGGGAGAATGTGCA 257  
Db 242 LeuLeuHisArgArgProHisHisGlyProPro----- 252  
QY 258 TATGATGGCTGATTACATCGACCTGGCGAAGTTCAAGTCCATCTACGTGGTCGACCTGTG 317  
Db 253 -----GlnArgAlaGlyArg-----HisArgArgArgLys----- 262  
QY 318 CCACTCGTGTGCGAGGTGGCCAAAGAGAGAGGCGCAAGCGCTGGAAGAATGTCCA 377  
Db 263 -----GluGlyArgArgArgGlyValGlnAspArg 274  
QY 378 GGTGTGGAGCGCGACCTTGCCTCAATTTGCGCCCTCGAGG-----CACGCGAC 428  
Db 275 HisArgGly-----ArgLeuLeuGlnProGlnGlyGlyProArgHisArgGln 292  
QY 429 GCTCATCACCTTCTCTACTCTGCTCAGATGATTCACCGTTCACCAACGTCTCGACCA 488  
Db 293 ArgHisValArgGlyLeuArgAlaArgGlyHisArgAspVal---ArgArgGlnArgPro 311  
QY 489 GGCTGTGCTGTACCTGCCAAGACGCGCTGGTGGGGTTCGCGACTTCTACGT----- 542  
Db 312 GlyArgProValGlyGlyProValArg-----ArgLeuLeuArgGlyHis 326  
QY 543 GAGCGCAAGTACGACCTGCCCTCGCCAGATGCCCTGTGTCGCGCGCTTCTTCTGCGG 602  
Db 327 GluArgGlnAlaArgValHisGlyProProAsp----- 337  
QY 603 ATCGATCTTCGACATCGACACATTGACATGGCGCGCGCGCTTACCTTGGAGCA 662  
Db 338 -----ProGlnAlaGluAlaProProGlyVal 346  
QY 663 GAAGCTGGAGCGGTGGGAGCAGAACACCGAGGTTCGATCCCTACGTGCGCGTGCCT 722  
Db 347 HisArgGlyArgGlyHisHisGlyAlaHisProGlyTrpGlnLeuLeuHisGluAla----- 365  
QY 723 GCGCGCCCCCTACTACGTGTGGATTGGCGCGCTGCCAGCGTTGGCCACGCGCTGCA 782  
Db 366 -----GlyGlnGluGlyGluArgAlaGlyPro 374  
QY 783 GGAGCGGTGGAGCGCGCGCATGTTCCGCCCGCACCTTCTGTACAGCGAGTGTGGA 842

375	AlaAlaGluAlaGlnAlaGlyGlnValArgAla-----ProHisValAlaAlaValAla	392
843	GGACCCCGAGCCGGATATGGAGGTGATGAGATCAACCCCAAGCAGCACGGTCTCACCT	902
393	GlyPro-----ProAspArgGlyHisProArgArgHisGlnValHisArg	407
903	GACTAGCGGGCTGCAATGCCCTGAACTGCTGTGTGTCAGGGGGCGCCAGTGTGTC	962
408	AlaArgGlyGlnLeuArgGluArgGlnPro-----	417
963	GTGGGACTCAACCCCGCAGCTCGCGCTTCTGGAGCTGAAGAAGTGGCCATTCA	1022
418	GlyHisArgArgProProArgGlnGlyAlaAlaArgArgGlnLeuProGlyHisProHis	437
1023	GCTGGAGTTTGAGCAGCTGTGGCAGCTGTTTCGGCCAGGCGCTGCACCCGCGATTGAG	1082
438	ArgArgValHisGlyGlnArgProProArgHisArgGlnHisArgGlnAlaHisValArg	457
1083	GCT-----GTACGAGAAGACTGGCGCCCTTCTCTCTCGCAAC	1122
458	AlaValLeuArgAlaArgGlnArgValLeuGlnGlnArgAlaHisLeuGlnProGlyArg	477
1122	CAGCCACA-----	1133
478	GlnProGlnProGlnProGlyLeuArgLeuGlnGlyHisArgAspArgHisGlyLeuLeu	497
1134	CTGTCTCAACGCGCTCTGTGTACTTCCAGCAGCGCTGTACTACAGCGGGCGCATGGCAA	1193
498	LeuLeuArgAlaProVal-----ProGlyGlnProHisHisGlnProArg	512
1194	GCTGTGTCTGGTGTCTGCAGTGCCTGCGCGCT-----GGTCTGGGACTGGGCAA	1241
513	AlaGluArgGlyAlaAlaGlnProGlyArgGluLeuProGlyProArgLeuGlyGlnGlu	532
1242	GACCGTCAGCGCTCTGCCAA-----CGCGCCCAACAATGCA-----	1277
533	AspArgArgGlyAspArgHisProGluAlaHisValValHisLeuHisArgGlyAlaVal	552
1278	-----GGAGCAGCGCG-----TCTGTGGGACAGCAACATGCT	1310
553	ProGlyArgGlyProAlaProProArgGlyGlyUHisGlnGlyValGlyGlnGluHisArg	572
1311	CATCCACTTCTGTGAAGAACGGCGCCCAAGCGCTGGTGTGCTTCTCAAGTTCGTGAG	1370
573	AspProGlyGlyGlnGluGlyAlaAsp-----HisGlu	583
1371	CCTGGTGTCTTCAAAGCGCGTGTGTGTGTCGGCGCGCGTCCGGGCAAGCAGTA	1430
584	Pro-----LeuGlyArgAlaLeu-----GlnArgProLeuGlnArgGlu	596
1431	CGCGCTGATCAAGCGGACGCAT-----CCCATTTAGAACTACATCGCGGCCACCATGA	1487
597	GlyAlaAspGln-----ArgHisArgProArgGlyArgValHis-----ValArgGly	612
1488	CGCGGTTCGGGAGAACTCGCACGTGCGCAAGCAAGCACTACTTCTTACAACTGCCTCAC	1547
613	GlyArgGlyGlnArgGlnProAlaAlaAspAlaGlu-----	624
1548	CGGCAAGTTCTTCGGCGCAAACTGCCCACTACTCGCGGAGCGCGCTTCGCCACCCT	1607
625	-----AlaAlaArg-----ArgAlaGlyGlyProArgPro	634
1608	CAAGAGTGGCTGTGGACAACCTGACGCTCTCCACCAACTTCTTCATGGAGGAGCTCAA	1667
635	GlnGlnArgArg-----ArgGlyAlaGly	642
1668	AGCGCGCACTTA-----CACCAAGGTGATTTCTGATGACCACCGTGAAGTGGCT	1715
643	AlaLeuHisValLeuGlnAspHisGlnValArgGlyGlyAlaProArgGlyAlaAlaPro	662
1716	-----GGATATGCCGTGGCCAAACGAGCTGCCAGTGCCT-----	1751
663	GlyGlyGlyValArgProArgGlyValArgArgGlyHisArgProArgGlyGluProAsp	682

```

Qy 1752 ---GGCCAGCAGCAGTTCGCCGG-----CGCATCGTCATCGCGCTCCGCGCTC 1799
      |||||
Db 683 ArgGlyGlnProValValProAlaValProLeuArgAlaArgGlyAlaArgLeuArgVal 702
      |||||
Qy 1800 CCTCAGCCCGCCCTACGCCGAGCTGATCCAGAAGCGCGCTTCAGCGTGGCTGCATCCG 1859
      |||||
Db 703 ProAsp-----ArgArgGluAlaGlnVal 710
      |||||
Qy 1860 CCSCGCCACTCAGGGGTACATGACCGCGTCAACATGTACAGCTCTTCTACATGCGCCG 1919
      |||||
Db 711 ProArgArgGlyValGlnGlnGlnValArgHisGlnProGlyGlnAlaArgGlyPro 730
      |||||
Qy 1920 -----CGGGAAGGG 1928
      |||||
Db 731 HisAlaArgValProGlnGly 737
      |||||

RESULT 14
US-11-182-016-20
; Sequence 20, Application US/11182016
; Publication No. US20060019294A1
; GENERAL INFORMATION:
; APPLICANT: SUGEN, INC.
; TITLE OF INVENTION: TYROSINE KINASE SUBSTRATE (TKS) PROTEINS
; FILE REFERENCE: 038602/0102
; CURRENT APPLICATION NUMBER: US/11/182,016
; CURRENT FILING DATE: 2005-07-15
; PRIOR APPLICATION NUMBER: US/09/958,359
; PRIOR FILING DATE: 2002-02-05
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 1717
; TYPE: PRT
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Grub
US-11-182-016-20

Alignment Scores:
Pred. No.: 3,1e-08
Score: 257.50
Percent Similarity: 29.6%
Best Local Similarity: 23.9%
Query Match: 7.1%
Gaps: 7
Dr. 51

```

QY 1636 -----CGGTCAAGTTGTCACACCGCCACCTCTTGAGGGTGG----- 1601  
Db SerProArgSerGlyCysProArgGlySerLeuGlyValAlaProSerArgGly 395  
QY 1600 -----CGA 1598  
Db CysGlyArgSerLeuArgGlyLeuProProGluAlaArgGlyAlaArgArgArg 415  
QY 1597 AGGCGCGCTCGCGAGGTAGTGG----- 1574  
Db ArgProGlyGlyArgThrThrTyrProSerGluAsnArgGlnArgLysGlnLysLys 435  
QY 1573 ---GGCAGTTGTCGGCA---GGAACTTCGGGTGAGGAGTGTGTAGTAGTGTCT 1520  
Db GluAlaSerCysArgSerArgGlySerProArgArgGlyGlnCys----- 450  
QY 1519 GCTTCGCGACGTGCGAGTTCTCCGCGACCGCTCCATGCTGCGCGAGTGTCTCA 1460  
Db ProThrGluProTrp-----GlyGln 457  
QY 1459 TGGGATGCGTCCGCTT-----TGATCAGCGCGTACTGCTTGGCCGCGACGC 1412  
Db GlyArgCysGlnProProArgSerProTrpGlnSerAlaLeuThrLysAlaGlnAlaSer 477  
QY 1411 CGC-----CGCGAAACCAACAGCACCGCGCTTGTGAAGAGCACCA 1373  
Db ArgMetProProGlyGlyGlyLeuArgGluAlaArgGlnAlaLysArgAlaGln 497  
QY 1372 GGCTCAGCACTTGACGACAGCACACACCGCGCTTGGCC-----CGT 1328  
Db AsnSerArgArgGluArgAlaSerAlaLeuSerLeuTrpAlaTyrArgArgGlyArg 517  
QY 1327 TCTTCACCAAGTGTAGTACGTGTGCTCCACAGCGCGTCTCTCCCTTCTGG 1268  
Db ArgGluArgAlaGly-----GlyAlaArgProAlaValTyr 529  
QY 1267 GCGGCTTGGCAGCGCTGTGAGTGTGCTTCCAGCAGCACCGCCAGCACGTGCA 1208  
Db GlyArgThrHis-----ArgProArgArgProProValHis 541  
QY 1207 GCACCCAGCAGCTTGCATCGCCGCTGTAGTACAGCGCGTGTGGAGTACCAGA 1148  
Db SerAsnThrSerAlaGlyAspCys-----AlaGlyArgLysArg 554  
QY 1147 GCGCTTGGACCAAGTGTGGCTGTTGCGCAGGAGGCGCGCTCTCTCTCGT 1088  
Db GlyGlnHisSerArg-----ArgGlyProCysSerLeuArg 566  
QY 1087 ACAGCTCTCAATGCGGGT-----GCACGCGCTCGCGCAACAGCT 1046  
Db LeuSerProCysSerLeuGlyLeuAspGlyIleTrpIleProHisProAspGlyArgGly 586  
QY 1045 GCCACAGTCTCAACTCCAGCTGCTGAATGGCA-----CCTTCTCAGTCCA--- 995  
Db GlyProGluTrpAlaSerSerAlaAspHisTyrProThrValProSerGlyAlaProThr 606  
QY 994 -----GAAGCGCGCGACT 983  
Db LeuProArgHisAlaGluHisAsnSerSerLeuProProLeuThrAlaGlnAlaSerThr 626  
QY 982 GCGCGGGTTGACGTCCACCGACACCACTCGCGCGCGCCCTCGCACCGAGGTTACAGG 923  
Db AspThrGlyAlaValArgProAlaGlyProSerSerGlyThrSerThrAlaSerSerThr 646  
QY 922 CAT---TGCACCGCGCTAGTACAGG-----TCAGCACCGTGTCTTGGGGTTGA 875  
Db HisSerCysLeuGluProThrSerGlyLeuArgArgSerProCysSerAlaAla 666  
QY 874 TCTCCATCACCTCCATAT----- 857  
Db AspSerHisSerProSerAsnThrLeuTrpIleSerGlyCysGlyAlaValArgGluSer 686  
QY 856 -----CCGCGTCCGGGTCT-----CCACG 836

Db GluLysSerGlyGlnAlaArgGlyAlaAlaValGlyValArgSerGlnGlyProLeu 706  
QY 835 ACTCGGTGTACAGNAGG---TGGCGCGGAACATGGCGCGCTCCACGCGCTCTCGT 779  
Db SerGlnSerLeuGlyArgAspThrProGlySerGlyLysAlaMetSerProValProArg 726  
QY 778 ---GCAGGCGCTGCGCAACCGTGGCGAGC---GGCCAATCCACACGTAGTAGGGCGC 725  
Db CysAlaGlyLeuGlyThrAlaGlyHisGlyAlaGlySerSerArgAlaArgGlyArg 746  
QY 724 GCA-----GCCACGCGACGTAGG----- 707  
Db GlyGlySerGlyAsnAlaGlnAlaThrAlaGluGlyAlaGlyArgSerProAlaAspGly 766  
QY 706 -----GGATCGAACCTGGGTGTCTCTCCACACGCGCTCCAGCTTCT 662  
Db ThrAlaGluGlyTrpGlyGlyHisProAsp---GluAlaAlaLeuHisSerGlnGln 785  
QY 661 GCTCCAGGTAGGCGC-----GGCGCTCGGGCGCGATGTCAATGT 623  
Db AlaGlyGlyProArgProSerTyrThrValSerGlySerGlyArgGlyHisSerProAla 805  
QY 622 TGT-----CGATGTGGAAGATCGATCCCGCAGAAACCGCGCGACCGGCGATCT 572  
Db CysAlaProLeuGlnProAlaArgAlaAlaAlaArgAlaAlaValProAlaAlaThr 825  
QY 571 GCGCGCAGG-----GCAGGTCTGACTTGC----- 548  
Db ProAlaGlyValAlaValAlaLeuGlyProArgGlyGlyAlaAlaGlyLysLeuCysTyr 845  
QY 547 -----CGCTCAGTAGAAGTCGG-----CAACGC 524  
Db AlaTrpGlyHisLeuValCysProAlaGlyAspArgAlaAlaIleProCysPheGlnArg 865  
QY 523 CCACAGCGCTTGTGGACAGGTACGAGCAAGCTGTGTCGATCGATCGTTGTGGAACGTG 464  
Db GlyProGlyAlaProGlyProGlyThrGlyGlyProGlySer----- 879  
QY 463 GAATCATCGTGAGCGAGTAGGAGAAGTGTAGCGGTGCGGTGC-----CCTCAGGG 410  
Db -----GlyGlyGluCysHisLeuProGluGly 888  
QY 409 GCGCAATTTGCAAGCGTCGCGCTCCACGACCTCGACAT-----TCTTCCAGC 362  
Db Ala-----GlyTyrLeuThrAlaAlaGlyAlaGly 898  
QY 361 CTTGGCGCTTGGCTTCTTGTGGCCACCTCGCACAGG-----AGTGGACA--- 314  
Db GluTrp-----ProProSerGlyProAlaAlaThrAlaLeuLeuProAlaGlyThrMet 916  
QY 313 -----GGTCGACACCGTAGATGGACTTGAACCTCGCCA---GGTCGATGT 272  
Db GlyGlyGlyLeuCysSerAlaGlyArgSerTrpAlaGlySerGlyGlyCysAlaGlyCys 936  
QY 271 AATCAGCCCATCATCATCGACATCTCCCGAGTGCACCCAC----- 233  
Db ThrGlyProAlaAlaGlyProArgAlaGlnCysArgHisLeuProGlyAspAlaGlyPro 956  
QY 232 -----CCAGGTCAACCCAGATGAGTTGAGCGCTCGGCCAGCGCGGCGACCAAGG 182  
Db GlyProGlyProGlyGlnPro-----SerSerProAlaArgMetGlyProLeuProGly 974  
QY 181 CAGCGCGCTGGGCGCGTAGAAGCTCTCCAGGCGAGCAG---CGTGTATCATCGCGCTTCT 125  
Db ProLeuProArgAlaArgGluGluAspProAlaThrProGlyArgGlyGlyGluProThr 994  
QY 124 TGCTGCCGAACCATATGCG----- 104  
Db GlyLeuProThrThrAlaGlyArgArgCysGlnGlnTrpArgGlyProValGlyAlaPro 1014  
QY 103 -----GCAGACGGTCAAGTTCATCTTCATGCTGC 74  
Db ----- 111



Search completed: March 14, 2006, 01:53:18  
Job time : 49 secs



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2006, 01:57:16 ; Search time 26 Seconds  
(without alignments)  
2398.018 Million cell updates/sec

Title: US-10-620-914-45

Perfect score: 3463

Sequence: 1 MSGRDRPASVTKNFSLE.....RVNYSFFYARRKGAKKDN 648

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 80:\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	326.5	9.4	416	2 A97614	hypothetical prote
2	326.5	9.4	416	2 AG2836	conserved hypothet
3	120	3.5	449	2 S07714	T64 protein precu
4	120	3.5	451	2 I50131	clusterin - quail
5	112.5	3.2	2285	1 G02434	DNA-directed DNA p
6	106.5	3.1	180	2 B69524	rRNA (adenine-N6) -
7	106	3.1	472	2 I59087	ISG-K54 - human
8	106	3.1	2039	2 S64540	probable calcium c
9	105	3.0	234	2 T48892	2-heptaprenyl-1,4-
10	104.5	3.0	664	2 F83376	conserved hypothet
11	102.5	3.0	465	2 AF1858	hypothetical prote
12	102.5	3.0	493	2 T09919	hypothetical prote
13	102	2.9	656	2 A48439	dnak-type molecula
14	101.5	2.9	663	2 E90689	cytochrome o ubiqu
15	101.5	2.9	663	2 B42226	bo-type ubiquinol
16	101.5	2.9	663	2 A85540	bo-type ubiquinol
17	99.5	2.9	663	2 AE0557	cytochrome o ubiq
18	98.5	2.8	552	2 E72283	alpha-galactosidas
19	98.5	2.8	673	2 T31610	hypothetical prote
20	98.5	2.8	1336	2 T02736	probable SCARECROW
21	98.5	2.8	2890	2 B71846	dna-directed RNA p
22	98	2.8	993	2 A10669	probable virulence
23	97.5	2.8	280	2 T36376	hypothetical prote
24	97.5	2.8	662	2 B75544	arginine decarboxy
25	97	2.8	239	2 AH2836	O-methyltransferas
26	97	2.8	239	2 B97614	hypothetical prote
27	97	2.8	363	2 G82376	recP protein VC001
28	97	2.8	366	2 T33885	hypothetical prote
29	96.5	2.8	152	2 AE0356	transposase, IS154

ALIGNMENTS

RESULT 1

A97614

hypothetical protein AGR\_C\_3843 [imported] - Agrobacterium tumefaciens (strain C58, Cerc

C;Species: Agrobacterium tumefaciens

C;Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 09-Jul-2004

C;Accession: A97614

R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001

A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens

A;Reference number: A97359; MUID:21608551; PMID:11743194

A;Accession: A97614

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-416 <KOR>

A;Cross-references: UNIPROT:Q8UDK6; UNIPARC:UPI00000D1DB1; GB:AE007869; PIDN:AAK87866.1;

C;Genetics:

A;Gene: AGR\_C\_3843

A;Map position: circular chromosome

QY	275	LYTQSWEDPEPMVEINPKDVLTLTSGGNALNLLVQAGQVVSVCNPAQSALEL	334
DB	44	VTFQWEDPEIDMEAMELGEGHRIIVTIGSGCNMLAYLSRNPASIDVDLNPHEIALNKL	103
QY	335	KKVAIQQL-EFEDVMQLFGEGVHPRIEELYEKKLAPFLSQTSHNFSK-----RLWYF	386
DB	104	KLAAPFHLPAHQDVVRHFGRAGTRNSVGVDFRFAEHLDAATTKAYWSKRTLSGRRRISVF	163
QY	387	QHGLYQQGNGKLCWVLOCLAVVLGKTVKRLANAPTEBQRRLWDSNMLHFVNGPK	446
DB	164	DRNIYRTGLLGRFICAGHIMARLHGKLT--EMAKTRTLDQEQPFDSKVAPLF-----DK	217
QY	447	PLV-MLFVKFVSLVLFNKAVLMFGGVPCKQY---ALIKADGPIPIENIARTWDGVAENS	502
DB	218	PVVRMLTKRKSLSL-----FGLGIPPRQVDELASLSDG--TVASVLKERLEKACNP	267
QY	503	HYRKQNYFYINCLTGKFLRDN---CPTYLREAAFLTKSGVVDNLVTSTNFFMEELK---	556
DB	268	PL-SDNYFAWQAFARRYPEPEHGALPAYLKPEYKIRNNTA-RVAVHHATYTELLSRKP	325
QY	557	ARTYTKVILMDHVDMLDMVANELAECLAKQVAPGGIVWRSAS-----LSPPYAEL	608
DB	326	ANGVDRIYLLDAQDWMTDVLNMLNSQISRTAASGARVIFRTAAEKSVIEGRISP-----	380
QY	609	IQKAGFDVTR-----CIRRATQ-GYMDRVNMYSSFFYARR	641
DB	381	-----DIRNQWYLEERSNELNAMDRAIYGGFHIYQR	413

30	96.5	2.8	152	2 AC0003	transposase, IS154
31	96.5	2.8	152	2 AE0195	transposase, IS154
32	96.5	2.8	152	2 AC0127	transposase, IS154
33	96.5	2.8	152	2 AE0349	transposase, IS154
34	96.5	2.8	152	2 AF0154	transposase, IS154
35	96.5	2.8	189	2 H70323	hypothetical prote
36	96.5	2.8	471	2 E83130	conserved hypothet
37	96	2.8	434	2 D72353	lipopolysaccharide
38	96	2.8	472	2 S29243	interstitial colla
39	95.5	2.8	426	2 B53652	rhmannosyltransfera
40	95.5	2.8	448	2 S45112	transcription fact
41	95.5	2.8	2890	2 P64669	DNA-directed RNA p
42	95	2.7	252	1 B48653	hypothetical prote
43	94.5	2.7	152	2 AF0012	transposase, IS154
44	94.5	2.7	169	2 T14928	transposase - Yers
45	94.5	2.7	2470	2 S57085	1-phosphatidylinos

QY 266 RP-----PM-----PPPTFLYTQ-----SWEDPEPDMEVMEIN-PKOTV-----L 299

Db 230 RELHPFFQHPMHGHRFLFQFLFEMTQHMULDGGHAWHEHPLGGPATRSRNFSTDMVCREI 289  
QY 300 TLTSGGCNALNLVQAGQVVSVC---NPAQSALELKKVAIQOLEPEFEDVWOLFGEVGH 356  
Db 290 RNSAGCLMRDCEKREILAVDCSQTDVQSQLE---QFEDALRL-----A 335  
QY 357 PRIEELYEKKLAPP-----LSQTSNPF-WSKRLMYFOHGLYYQGMGKLCWVLOC 405  
Db 336 ERFTRRYDOLLSAFOAEMLNTSLLDQLNRQFGWVSR-----GNLTQNDG---FLQV 386  
QY 406 LAVVLGLGKTVKELANAPTMEEQRRLWDSNMLHFVNGPKPLV-----WLFVKVFSLV 459  
Db 387 TTV-----FSKTPNLEDSAPADQTVTQVLFDSPLSLTVPGDISWDDPRFMEIV 436

RESULT 5  
G02434  
DNA-directed DNA polymerase (EC 2.7.7.7) epsilon catalytic chain A [validated] - human  
C:Species: Homo sapiens (man)  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 20-Apr-2000  
C:Accession: G02434; A46692  
R:Asahara, H.; Goldsmith, J.S.; Lee, E.; Linn, S.  
submitted to the EMBL Data Library, February 1996  
A:Reference number: H01252  
A:Accession: G02434  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-2285 <ASA>  
A:Cross-references: UNIPARC:UPI000016A2D7; EMBL:U49356; NID:g1206034; PIDN:AAA90924.1; F  
R:Kest, T.; Frantti, H.; Syvaioja, J.E.  
J. Biol. Chem. 268, 10238-10245, 1993  
A:Title: Molecular cloning of the cDNA for the catalytic subunit of human DNA polymerase  
A:Reference number: A46692; MUID:93252906; PMID:8486689  
A:Accession: A46692  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-68,96-441,'T',444,'AGS',448-1081,'NA',1084-1297,'E',1299-1371,'P',1373-151  
A:Cross-references: UNIPARC:UPI000016A096; EMBL:S60080; GB:L09561; NID:g303156; PIDN:AA  
A:Experimental source: HeLa cells  
A:Note: sequence inconsistent with the nucleotide translation  
A:Note: sequence extracted from NCBI backbone (NCBIN:131270, NCBIP:131271)  
C:Genetics:  
A:Gene: GDB:POLE  
A:Cross-references: GDB:129548; OMIM:174762  
A:Map position: 12q24.3-12q24.3  
C:Superfamily: DNA-directed DNA polymerase II  
C:Keywords: DNA binding; nucleotidyltransferase

Query Match 3.2%; Score 112.5; DB 1; Length 2285;  
Best Local Similarity 19.4%; Pred. No. 4.9;  
Matches 135; Conservative 83; Mismatches 199; Indels 279; Gaps 34;  
QY 107 HSLCEVAKKAKAGKHNQVVEADACQAPPEGATLTFFYS-----150  
Db 422 HNL-----KAAASKGLGDPVELDPEDCMRA-TEQPQTATYSVSDAVATYILMYKVVHPF 476  
QY 151-----LTMIPPH-NVI-----DOACSVLSODGLGVAD 178  
Db 477 IFALCTIIPMEPDEVLRKSGTLCCEALLMVQAFHANIIPNKQEQBFNKLTDDGHVLDSE 536  
QY 179 FTVSG-----KYDLPRLQMPMSRRFFWRSIPDIDNIIGPERRAYLQKLERVWE 228  
Db 537 TVVGHVGALESQVFRSDIPCR-----FRMNPAAFDLLQVKEK-----575  
QY 229 QNTQGSIPVPLRAPPYVWIGRLPSVGHALHE-----262  
Db 576-----TLRALBEEKVPVEQVNTFNEVCEDEIKSKLASLK 610  
QY 263-----RVERP-----PMFPFTLYTQSWEDPEPDMWEINPKDVTLTLSGCN---307  
Db 611 DVPSRIECPLIYHLOVGAMYPNLTNR-----LOPSANVDEATCAACDFNK 657

QY 308 -----ALNLLVQAGQVVSVCNCPAQSALELKKVAIQOLEFEDVWOLFGEVGHPIREEL 362  
Db 658 PGANCQRKMAWQWRGSEFM-----PASRS--EYHQI-Q-HOLESEKFPPLPEGPAPAFHEL 709  
QY 363 -----YEK-KLAPP-----LSQTSNPFWSKRLMYFOHGLYYQGG 395  
Db 710 SREOAKYEKRRLADYCRKAYKIHITKVEERLTTTCORENSFYVDTVAFRRDRRYEFKG 769  
QY 396 MGKLCWVLCQAVVLGIGTKTVKRLANAPTMEEQRRLWDSNMLIH-----FVKNGP 445  
Db 770 LHKVKKKLSAAVEGDAAEVKCKN---MEV---LYDSLQLAHKCILNSFYGVWRKGA 823  
QY 446 KELVWLFFVKVSLVFNKAVLWFGGVGPKQKVALIKADGIPENYIARTWDG---VAENS 502  
Db 824 R---MYSMEMAGIVCFT-----GANIITQARELIQIGRPLE---LDTDGIWCVLPS 870  
QY 503 HVRKONYFYVNCITGKFLRDNCPITYLREAPATLKSGVVDNLTVSTNF-----FMBELKAR 558  
Db 871 F---PENFVFKT-----TNVKPKVTISYPGAMLNIMVKEGFTNDQIQLAEP 916  
QY 559 TYTKVILMDHVDWLD-----MPVANLEAELAKQVAPGGIVIVRSASLSPPYAELI 609  
Db 917 SLTYVTRSENSIFPEVDGPLYAMILPASKEGKCLKRYA-----VFNEDGSL---AEL- 967  
QY 610 QKAGFDVRCIRATQGYMDRVNMY-SSFYMARCKGA 644  
Db 968 --KGFEVK-----RRGQLIKIFQSSVFEAFKGS 996

RESULT 6  
B69524  
rRNA (adenine-N6)-methyltransferase homolog - Archaeoglobus fulgidus  
C:Species: Archaeoglobus fulgidus  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 12-Jul-2004  
C:Accession: B69524  
A:Klenk, H.P.; Clayton, R.A.; Tomb, J.P.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson,  
; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;  
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
Nature 390, 364-370, 1997  
A:Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.  
Smith, H.O.; Woese, C.R.; Venter, J.C.  
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo  
A:Reference number: A69250; MUID:98049343; PMID:9389475  
A:Accession: B69524  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-180 <KLE>  
A:Cross-references: UNIPROT:O28089; UNIPARC:UPI0000056A3B; GB:AE000954; GB:AE000782; NID  
C:Superfamily: spore germination protein C2

Query Match 3.1%; Score 106.5; DB 2; Length 180;  
Best Local Similarity 22.3%; Pred. No. 0.43;  
Matches 48; Conservative 38; Mismatches 64; Indels 65; Gaps 14;  
QY 34 LRHMFGSKKGD--DHAARLESFYGPQAAFAARLAERSNLIVDLGGGTGENVMADY 91  
Db 1 MKHK-FDPKKAHLDSEWRKIFPPEKVFETESLRPKQVLPF-DVGAGTG-----Y 50  
QY 92 --IDLAK-FKSYVVDLCHSLCEVAKKAKAGKHNQVVEADACQAPPEGATLTITS 148  
Db 51 LTIPLARVFKKYAYVEISEMAEVLRLRVEEGLNIGIIVSE---KPE-----VDPR 101  
QY 149 YSLTWIPPHNVIDQACSVLSODGLGVADFYVSGKYDLPRLQMPMSRR-----PFWR 202  
Db 102 VDVL---FSNVLHE-----MDNPEYLEWASADYVVAEWK 137  
QY 203 IPDIDNIDIGP--ERRAVLEQ-----KLERVWEQN 230  
Db 138 ----EKTEFGPPVEERLSLEENLSKMKLVKSEN 168

RESULT 7  
159087



Db 732 SRKTGYVAR 740  
RESULT 9  
T48892  
2-heptaprenyl-1,4-naphthoquinone methyltransferase [validated] - Bacillus stearothermophilus  
C:Species: Bacillus stearothermophilus  
C>Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 09-Jul-2004  
C:Accession: T48892  
R:Koike-Takeshita, A.; Koyama, T.; Ogura, K.  
J. Biol. Chem. 272, 12380-12383, 1997  
A:Title: Identification of a novel gene cluster participating in menaquinone (vitamin K2) synthesis of Bacillus stearothermophilus.  
A:Reference number: Z24853  
A:Accession: T48892  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-234 <K01>  
A:Cross-references: UNIPROT:O86169; UNIPARC:UPI0000012EF03; EMBL:D87054; PIDN:BAA32500.1  
C:Genetics:  
A:Gene: meng  
C:Function:  
A:Description: catalyzes the transfer of a methyl group from 2-heptaprenyl to 1,4-naphthoquinone  
A:Pathway: methaquinone biosynthesis  
C:Superfamily: spore germination protein C2; bioC homology  
C:Keywords: methyltransferase  
Query Match 3.0%; Score 105; DB 2; Length 234;  
Best Local Similarity 23.5%; Pred. No. 0.82;  
Matches 24; Conservative 24; Mismatches 42; Indels 12; Gaps 2;  
QY 100 IYVDLCHSLCEVAKKAKGKGNVQVVEADACQAPPEGTATLITFSYSLTMIPFFHN 159  
DB 75 VYGLDFSENMLKVGQKVKARGLHNKLIHGNAMQLPFPDPSFDYVTIGFGLRNVDYWT 134  
QY 160 VLDQ-----ACSVLSQDLGVAD-FYVSGYDLPL 189  
DB 135 VLKEMHRVTKPGGITVCLTSTQPTLFGFRQLYFFFRFIMPL 176  
RESULT 10  
F83376  
conserved hypothetical protein PA2151 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 09-Jul-2004  
C:Accession: F83376  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Boman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen  
A:Reference number: A82950; MUID:20437337; PMID:10984043  
A:Accession: F83376  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-664 <STO>  
A:Cross-references: UNIPROT:Q911W4; UNIPARC:UPI000000C55D5; GB:AE004642; GB:AE004091; NID  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA2151  
Query Match 3.0%; Score 104.5; DB 2; Length 664;  
Best Local Similarity 19.4%; Pred. No. 3.8;  
Matches 93; Conservative 55; Mismatches 170; Indels 161; Gaps 25;  
QY 42 KGDDHAARLESFYGQAAAPAAARLARN-----LIWVDLGGGTGGTGVNDMMADYI 92  
DB 132 REGDELLQCAERGGPEIAAACAPLAERLQACQSVVEERVALM--LAAQTGELLRLVGP 189  
QY 93 DLAKFSIYVDLCHSLCEVAKKAKGKGNV--QVVEADACQAPPEGT-----A 142  
DB 190 HLVRSEYFPV-----EVERPLARFASWYELFPRSEGDPTP-----HGTDFDVIIRLP 237

QY 143 TLITFSYSLTMIPFFHN-----IDQACSYL--SQDG-----172  
DB 238 QIAAMGFVLYPPIHPICRTHRGKNNSLRAEAGDPGSPYAIGSEGGHEAHPGLGDR 237  
QY 173 -----LVGV-----ADFYVSGKYDLP-LRQMP-WRRRPFWR-----SI 203  
DB 298 EDRRLLVAVRHGHEALDAIFAIQSPDPHWRHFGW---FAWRPDGSLRYAENPKKY 354  
QY 204 FIDNIDIGFERRAYLEQLERVWEQNTQGSIPYV-----PWLR-APYVYVI--249  
DB 355 EDVNVDF-----YAEQALPSLWEALRDVLGWVEQGVTLFRVDNPHTKPLPFWEMLIA 408  
QY 250 -----GRLPVSGHALHERVERPPM-----PPTFLY-----TQSW 280  
DB 409 EVRGRHPQV--IFLSEAFTRPAMMARLGKVGFSQSYTYFTWRNDKQSLAIFYAELNQP 466  
QY 281 ED---PEPDMVEVMEINPKDVTLTSTGGCNALNLLVQAG-----QVSVVDCNPAQS 329  
DB 467 RDCYRPNFVTPDINPWLQSRGPPGLIRALATMGSLGWMYSGFELCEAAALPGKE 526  
QY 330 ALLELKVAIQOLEFEDVWQLFEGEVHPRIELEYKK--LAPPLSQTSNFWMSKRLMYF 386  
DB 527 EYLDSEKYLPRDYOAPGNIVAE--IARLNIRRENPAALQTHLGFQAYNAWNRILYF 583  
RESULT 11  
AFI858  
hypothetical protein all0415 [imported] - Nostoc sp. (strain PCC 7120)  
C:Species: Nostoc sp. PCC 7120  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C>Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Jul-2004  
C:Accession: AFI858  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Kaneko, T.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AFI858  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-465 <KUR>  
A:Cross-references: UNIPROT:Q8YZP2; UNIPARC:UPI000000CDD3B; GB:BA000019; PIDN:BA872373.1;  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: all0415  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE29.06c  
Query Match 3.0%; Score 102.5; DB 2; Length 465;  
Best Local Similarity 19.7%; Pred. No. 3.4;  
Matches 75; Conservative 50; Mismatches 150; Indels 105; Gaps 18;  
QY 38 WFGSKGDDHAARLESFYGQAAAPAAARLARNL---IWDVLG-----GGTGENVDM 87  
DB 109 WYNNYSSVDYVTPRPMYPADGKPLDIYSQENLRSRLQIDLQGPLFNFPGNTSIRS 168  
QY 88 MADYIDLAKFKS-----IYVDLCHSLCEVAKKAK-ARGKGNVQVVEADACQFAP 137  
DB 169 TQWIADSAKFLDGRCDPTLTLYVLPHLDYCLQKYGVDITKIAKDLQIDAICSDLIQFYE 228  
QY 138 PEGTATLITFSYSLTMI-PPFH-----NVIDQACSYLSQDLGVVA 177  
DB 229 NFGAQVILVSEYGITSVSQPIHNRVLRHGLLTIREELGRELLDAGAKAFADVHQVA 288  
QY 178 DFVSGKYDLPURQMPWSRRFFRSIFDIDNID----IGPERRAYLEOKLERVWEQNTQ 232  
DB 289 HYVNDPYIP--QV-----RSL--LENIDGADVLDQKSYTHLNHRSAGELIAV 336  
QY 233 GSIPYVPMIRAPYVYVWIG--RLPSVGHALHERVERPPMFPPTFLYTQSWEDPE---PDM 287  
DB 337 SQ----PDAWFTYVYVLDLDRRAPDFAKTVD---IHRKPGYDVPVELFL---DPEIKLPQV 385  
QY 288 EV-----MEINPKDVTLTSTGGCNALNLLVQAGQVSVVDCNPAQSALLE 333

Db 386 KIATKLQKKGRFLYLMIDIPDLAELVKSHGC-----LPPSSQSPFLLI 430  
QY 334 LKKVAI---QQLEFEDVWQL 350  
Db 431 TQSHLFDSTAISATDVYQL 450

RESULT 12  
T09919  
hypothetical protein T16L4.70 - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 09-Jul-2004  
A;Accession: T09919  
R;Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.  
submitted to the Protein Sequence Database, June 1999  
A;Reference number: Z16897  
A;Accession: T09919  
A;Molecule type: DNA  
A;Residues: 1-493 <BEV>  
A;Cross-references: UNIPROT:Q9SU89; UNIPARC:UPI000000AC268; EMBL:AL079344; GSPDB:GN00062;  
A;Experimental source: cultivar Columbia; BAC clone T16L4  
C;Genetics:  
A;Gene: ATSP:T16L4.70  
A;Map position: 4  
A;Introns: 65/2  
C;Superfamily: Arabidopsis thaliana hypothetical protein T16L4.70

Query Match 3.0%; Score 102.5; DB 2; Length 493;  
Best Local Similarity 20.4%; Pred. No. 3.7;  
Matches 102; Conservative 65; Mismatches 171; Indels 161; Gaps 24;

QY 176 VADPYSGKDYDLPLRQMPSPRRFWSIFIDINIDIGPERRAYLEQKLERVWEQNTQCSI 235  
Db 111 VVDFWQVORDAHLNLDRIIP-KPKFEWISNLDLSAD-----EDSMKQVFD---SV 155  
QY 236 PYVPWLRAPYVYVWIGRLPSVGHALHEERVPMPPT-----FLYTQSW----- 280  
Db 156 P--DWLSE-----KLGSAGTIL-----PWLPSVCDVDSEMLVVDVSWNGKEITQI 198  
QY 281 -EDPEPDM-----VMEINPKDTVLTLTSGGCNALNLVQ-----GAGQVVSVDNCPAQSALL 332  
Db 199 KEDMEEDQREVVHNDHTMTVGLQAHDEMAVNLRAQISNFESTSEVLTIL-CNDIRRLCL 257  
QY 333 ELKVAIQLEFEDVWQLFEGVHPRIEELYEKLAFLSQTSHNFSKELWLFQHGLEY 392  
Db 258 EKGDSLRVLALTEPMNA-----DDETAVALLSLHN----- 289  
QY 393 QGGMGKLCWVLCQCLAVVLGLGKTVKRLANAPTMBEQRRLWDSNNMLHFVKNGPKPLVWLF 452  
Db 290 -----GSEEBELGFWSHVLCIV-----L 308  
QY 453 VKFVSL-----VLFNKAVLWFGGVGPKQYALI-----KADGPIPIENYIARTWDGV-AE 500  
Db 309 PKFLSLEKSASRVLSMTSTIEFKIHQRAAEYALVPLLRKEG--INNFCIEVISRVLKE 366  
QY 501 NSHVKQNYFYNNCLTGKFLRDNCPYLRNAFATL---KGVVDNLTVTNPFWEELXA 557  
Db 367 CLHLGQISGFCQKLLCGR-----TQERRFMFLPCHRDLTSDLTWNEYLFIPLFONI 417  
QY 558 RYTKVILMDHVDWMDMPVANELAECLAKVAPGGIVWRSASISPPYA-----ELIQKAG 613  
Db 418 LIHGIPISQSDVCLVSKV-QELAERYSKSLKFGNFIHFTAKCAPMLOAHKYLQILB--- 473  
QY 614 FDVRCIRR-ATQGYMDRN 631  
Db 474 -SVKCTNSLVTKSILSKIN 491

## RESULT 13

A48439  
dnak-type molecular chaperone Hsp70 - Entamoeba histolytica  
N;Alternate names: heat shock protein Hsp70

## RESULT 14

E90689  
cytochrome o ubiquinol oxidase subunit I [imported] - Escherichia coli (strain O157:H7,  
C;Species: Escherichia coli  
C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 09-Jul-2004

C;Species: Entamoeba histolytica  
C;Date: 01-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 31-Dec-2004  
C;Accession: A48439  
R;Ortner, S.; Plaimauer, B.; Binder, M.; Wiedermann, G.; Scheiner, O.; Duchene, M.  
Mol. Biochem. Parasitol. 54, 175-183, 1992  
A;Title: Humoral immune response against a 70-kilodalton heat shock protein of Entamoeba  
A;Reference number: A48439; MUID:93063033; PMID:1435858  
A;Accession: A48439  
A;Status: preliminary  
A;Molecule type: nucleic acid  
A;Residues: 1-656 <ORT>  
A;Cross-references: UNIPROT:Q24842; UNIPARC:UPI0000078A72; GB:M84652; NID:g158955; PIDN:~  
A;Experimental source: SFL-3, trophozoites  
A;Note: sequence extracted from NCBI backbone (NCBIN:117218, NCBIP:117219)  
C;Function:  
A;Description: involved in protein folding and assembling/disassembling of protein complex  
C;Superfamily: bcr protein  
C;Keywords: ATP; molecular chaperone

Query Match 2.9%; Score 102; DB 2; Length 656;

Best Local Similarity 19.0%; Pred. No. 6;  
Matches 121; Conservative 64; Mismatches 175; Indels 276; Gaps 27;

QY 57 POAAFAARLAERS-----NLIWVDLGGGT-----GENVDM 87  
Db 179 PTAALAYGLDKKSDREKNVLPDLGGGTDFVSLAIDDGVEFKASNGDTHLGGEDFDN 238  
QY 88 MADYIDLAKFKSYVVDL-----CHSLCEVAKKAKAKGKXVQVVEADACQAPP 138  
Db 239 RLNVHFIABPKRYKKDISGNARAVRLRTACRAKRTLSAATANIEV----- 287  
QY 139 EGTATLITFSYSLTMTIPPFHNVIDQACSYLSQDLGVADFYVSGKYDLPLRQMPSSRRF 198  
Db 288 -----DQLFDGIDFYTS----- 299  
QY 199 FWSIFIDINID-----IGERRAYLEQKLERVWEQNTQGISPIYVWLRAPYVWIG--- 250  
Db 300 ITRARFEELNIDLFKSTIGFVERVLQDAKLDK-----GSIDDV-----VLGGST 344  
QY 251 RLPSVGHALHEERVPMPPTFLYTQSWEDPEPMVEINPKDTV-----LT 300  
Db 345 RIPKVVQLLOD-----PFGKPEPN---KSNPDEAVAYGAQAAILT 384  
QY 301 LTSGGCNALNLVQAGQVVSVD--CNPAQSALLELKKVAIQLEFEDVWQLFEGVHPRI 359  
Db 385 GTGKATEDVLLDVAPLTLGIETAGGVMTALI-----PRN 420  
QY 360 EELYEKKLAPFLSQTSHNFSKELWYFQHGLEYQGGMGKLCWVLCQLAVVLGLGKTVKRL 419  
Db 421 STIPAKKSQVFSYADN-----QPGVLIVQFEGEASMTNHC----- 456  
QY 420 ANAPTMBEQRRLWDSNNMLHFVKNG--PKPLVWLVKFVSLVLFNKAVLWFGGVGPKQY 477  
Db 457 -----NLLGKFELTGPPAP-----RGVPQIEV 479  
QY 478 AL-IKADGPIPIENYIARTWDGVNSHVRKQNTFYNNCLTGKFLRDNCPYLRNA-AFAT 535  
Db 480 TPDIDANGI-----LNVSADKTTGKKNTITNDKGRLSKEQIDRWVAEAEFKA 530  
QY 536 -----LKSGV-----VNLTIVS-----TNPFWEELKARTYTKV--ILMDHVDWLD--MPVA 577  
Db 531 EDDKMQRVEAKNKLENFCYSVKNTLSEQPADKIAAEDKTTIENIVKETLWDINDNNQNAS 590  
QY 578 NELAECLAKVAPGGIVWRSASISPPYAEILQKAG 613  
Db 591 TEEYDNKQKEV-----EGKVPQIFTKLYQOAG 617



C;Accession: E90589  
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A;Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genome  
A;Reference number: A9629; MUID:21156231; PMID:11258796  
A;Accession: E90589  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-663 <HAY>  
A;Cross-references: UNIPROT:P18401; UNIPARC:UPI0000111624; GB:BA000007; PIDN:BA033908.1;  
A;Experimental source: Strain O157:H7, substrain RMD 050952  
C;Genetics:  
A;Gene: Ec80485  
C;Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology  
C;Keywords: copper; electron transfer; heme; iron; magnesium; membrane-associated complex  
F;106-421/Binding site: heme a iron (His) (axial ligands) #status predicted  
F;284-288/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted  
F;288/Binding site: oxygen (Tyr) #status predicted

Query Match 2.9%; Score 101.5; DB 2; Length 663;  
Best Local Similarity 22.2%; Pred. No. 6.7;  
Matches 63; Conservative 35; Mismatches 95; Indels 91; Gaps 15;

QY 235 IPYVPLRAPPYVWIGRLPSVGHALHEERVPMPPTFLYTQSWEDPPDMEVMEINP 294  
Db 179 IEYSPGVGDYVWISLQSLGIGTTLGIN-----FVITLKRA-----PGMTMFKM-P 226

QY 295 KDTVLTITSGCNALLVQAGQVSDCNPAQSALELKKVAIQOLEPED----- 346  
Db 227 VFTWASLCA-----NVLIASFPILTVT-----ALLTLDRLVGLTHFTNDMGGNNMMY 275

QY 347 ---VMQFGEVHPRIEELYEKKLAPP--LSQTSNFWSKRLMYFOHGLYQGGMGKLCW 401  
Db 276 INLIW-AWG---HP---EVYILLPVGFVSEIAATFSRKLFP-----GYTSLW 318

QY 402 VLOCLAV-----LGLKTVKRLANATMBEQRRLWDSNMLIHFKVNGPKPLVW 450  
Db 319 ATVCITVLSFIVLHFFHTMGAGANVNAFFGITM-----IIAIPGVKIFNW 366

QY 451 LFKVFSVLFPNKAFLW-----FGG-----GVPGKQYAL 479  
Db 367 LFTWYQGRIVFHSAMLWTIGFIVTFSVGGMTGVLAVPGADPVL 410

RESULT 15  
B42226  
bo-type ubiquinol oxidase (EC 1.10.3.-) chain I - *Escherichia coli* (strain K-12)  
N;Alternate names: cytochrome bo chain I  
C;Species: *Escherichia coli*  
C;Date: 19-Jun-1992 #sequence revision 19-Jun-1992 #text\_change 09-Jul-2004  
C;Accession: B42226; G64772; JC5635  
R;Chopuri, V.; Lemieux, L.; Au, D.C.T.; Gennis, R.B.  
J. Biol. Chem. 265, 11185-11192, 1990  
A;Title: The sequence of the cyo operon indicates substantial structural similarities be  
A;Reference number: A42226; MUID:90293062; PMID:2162835  
A;Accession: B42226  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-663 <CHE>  
A;Cross-references: UNIPROT:P18401; UNIPARC:UPI0000111624; GB:J05492; NID:g145651; PIDN:  
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co  
A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A;Title: The complete genome sequence of *Escherichia coli* K-12.  
A;Reference number: A64720; MUID:97426617; PMID:9278503  
A;Accession: G64772  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-663 <BLAT>  
A;Cross-references: UNIPARC:UPI0000111624; GB:AE000149; GB:U00096; NID:g1786628; PIDN:AP  
A;Experimental source: strain K-12, substrain MG1655

R;Kawasaki, M.; Mogi, T.; Anraku, Y.  
J. Biochem. 122, 422-429, 1997  
A;Title: Substitutions of charged amino acid residues conserved in subunit I perturb the  
A;Reference number: JC5635; MUID:98021083; PMID:9378723  
A;Accession: JC5635  
A;Molecule type: protein  
A;Residues: 1-55, DI', 58-300, AI', 303-330, LW', 333-510, G', 512-663 <KAW>  
A;Cross-references: UNIPARC:UPI0000175173  
C;Comment: This protein binds all the redox metal centers, low-spin heme b, high-spin heme  
C;Genetics:  
A;Gene: cyoB  
C;Complex: heterooligomer; the cyoABCDE gene products are required for bo-type ubiquinol  
ubunits II, III and IV may be required for the assembly of the metal centers in subunit I  
C;Function:  
A;Description: the cytochrome o complex catalyzes the oxidation of ubiquinol to ubiquinone  
C;Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology  
C;Keywords: chromoprotein; copper; electron transfer; heme; iron; magnesium; membrane-as  
F;16-32/Domain: transmembrane #status predicted <TM1>  
F;53-501/Domain: cytochrome-c oxidase chain I homology <CO1>  
F;57-73/Domain: transmembrane #status predicted <TM2>  
F;107-123/Domain: transmembrane #status predicted <TM3>  
F;146-162/Domain: transmembrane #status predicted <TM4>  
F;237-253/Domain: transmembrane #status predicted <TM5>  
F;287-303/Domain: transmembrane #status predicted <TM6>  
F;316-332/Domain: transmembrane #status predicted <TM7>  
F;347-363/Domain: transmembrane #status predicted <TM8>  
F;387-403/Domain: transmembrane #status predicted <TM9>  
F;414-430/Domain: transmembrane #status predicted <TM10>  
F;461-477/Domain: transmembrane #status predicted <TM11>  
F;498-514/Domain: transmembrane #status predicted <TM12>  
F;591-607/Domain: transmembrane #status predicted <TM13>  
F;610-626/Domain: transmembrane #status predicted <TM14>  
F;106-421/Binding site: heme b iron (His) (axial ligands) #status predicted  
F;284-333/Binding site: copper (His) #status predicted  
F;284-288/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted  
F;288/Binding site: oxygen (Tyr) #status predicted  
F;419/Binding site: heme o iron (His) (axial ligand) #status predicted

Query Match 2.9%; Score 101.5; DB 2; Length 663;  
Best Local Similarity 22.2%; Pred. No. 6.7;  
Matches 63; Conservative 35; Mismatches 95; Indels 91; Gaps 15;

QY 235 IPYVPLRAPPYVWIGRLPSVGHALHEERVPMPPTFLYTQSWEDPPDMEVMEINP 294  
Db 179 IEYSPGVGDYVWISLQSLGIGTTLGIN-----FVITLKRA-----PGMTMFKM-P 226

QY 295 KDTVLTITSGCNALLVQAGQVSDCNPAQSALELKKVAIQOLEPED----- 346  
Db 227 VFTWASLCA-----NVLIASFPILTVT-----ALLTLDRLVGLTHFTNDMGGNNMMY 275

QY 347 ---VMQFGEVHPRIEELYEKKLAPP--LSQTSNFWSKRLMYFOHGLYQGGMGKLCW 401  
Db 276 INLIW-AWG---HP---EVYILLPVGFVSEIAATFSRKLFP-----GYTSLW 318

QY 402 VLOCLAV-----LGLKTVKRLANATMBEQRRLWDSNMLIHFKVNGPKPLVW 450  
Db 319 ATVCITVLSFIVLHFFHTMGAGANVNAFFGITM-----IIAIPGVKIFNW 366

QY 451 LFKVFSVLFPNKAFLW-----FGG-----GVPGKQYAL 479  
Db 367 LFTWYQGRIVFHSAMLWTIGFIVTFSVGGMTGVLAVPGADPVL 410

Search completed: March 14, 2006, 01:58:35  
Job time : 28 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2006, 01:58:16 ; Search time 31 Seconds  
(without alignments)  
1728.188 Million cell updates/sec

Title: US-10-620-914-45

Perfect score: 3463

Sequence: 1 MSGRDRPASVTKNFKNSLE.....RVNYSFFYMARKGAKDN 648

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /cgn2\_6/prodata/1/iaa/5\_COMB.pep.\*
- 2: /cgn2\_6/prodata/1/iaa/6\_COMB.pep.\*
- 3: /cgn2\_6/prodata/1/iaa/H\_COMB.pep.\*
- 4: /cgn2\_6/prodata/1/iaa/PCTUS\_COMB.pep.\*
- 5: /cgn2\_6/prodata/1/iaa/RE\_COMB.pep.\*
- 6: /cgn2\_6/prodata/1/iaa/backfilese1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	352	10.2	310	2	US-09-248-796A-21233
2	158	4.6	222	2	US-09-248-796A-26335
3	121	3.5	212	2	US-09-107-532A-7254
4	118	3.4	329	2	US-09-902-540-16360
5	111	3.2	195	2	US-09-583-110-4761
6	111	3.2	198	2	US-09-107-433-4622
7	108.5	3.1	706	2	US-09-252-991A-24766
8	107	3.1	255	2	US-09-667-373-2
9	107	3.1	255	2	US-10-305-413-2
10	104.5	3.0	677	2	US-09-489-039A-13088
11	103.5	3.0	1070	2	US-09-107-532A-3716
12	101.5	2.9	663	2	US-09-711-164-332
13	99.5	2.9	471	2	US-08-994-689C-1
14	98.5	2.8	555	2	US-09-134-078-24
15	98.5	2.8	2636	2	US-09-252-991A-25753
16	97.5	2.8	464	2	US-09-134-001C-3488
17	97.5	2.8	471	2	US-08-994-689C-21
18	97	2.8	2209	2	US-10-017-754-1903
19	96.5	2.8	555	2	US-09-252-991A-28313
20	96.5	2.8	2470	2	US-08-265-967C-2
21	96.5	2.8	2470	2	US-08-305-790B-3
22	96	2.8	717	2	US-09-248-796A-18993
23	95.5	2.8	459	2	US-09-949-016-7329
24	94	2.7	480	2	US-09-438-185A-934
25	93.5	2.7	445	2	US-09-252-991A-17629
26	93.5	2.7	471	2	US-09-391-104-25
27	93.5	2.7	475	2	US-09-248-796A-15861

28	93.5	2.7	480	2	US-09-949-016-10560	Sequence 10560, A
29	93.5	2.7	590	2	US-09-489-039A-10743	Sequence 10743, A
30	92	2.7	574	2	US-09-605-703B-254	Sequence 254, App
31	92	2.7	1036	2	US-09-489-039A-10266	Sequence 10266, A
32	91.5	2.6	341	2	US-09-538-092-634	Sequence 634, App
33	91.5	2.6	681	2	US-09-248-796A-16367	Sequence 16367, A
34	91	2.6	749	2	US-10-104-047-2622	Sequence 2622, App
35	90.5	2.6	504	1	US-08-645-900A-1	Sequence 1, Appli
36	90.5	2.6	504	1	US-08-882-238A-1	Sequence 1, Appli
37	90.5	2.6	504	1	US-08-667-790A-1	Sequence 1, Appli
38	90.5	2.6	504	2	US-09-220-459-1	Sequence 1, Appli
39	90.5	2.6	504	2	US-08-938-669A-32	Sequence 32, Appli
40	90.5	2.6	504	2	US-08-546-568B-1	Sequence 1, Appli
41	90.5	2.6	504	2	US-08-822-999-3	Sequence 3, Appli
42	90.5	2.6	504	2	US-09-056-285A-8	Sequence 8, Appli
43	90.5	2.6	504	2	US-09-306-828-32	Sequence 32, Appli
44	90.5	2.6	504	2	US-09-952-464A-8	Sequence 8, Appli
45	90	2.6	500	2	US-09-543-681A-6847	Sequence 6847, Ap

ALIGNMENTS

RESULT 1

US-09-248-796A-21233

; Sequence 21233, Application US/09248796A

; Patent No. 6747137

; GENERAL INFORMATION:

; APPLICANT: Keith Weinstock et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA-ALBICANS

; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.132

; CURRENT APPLICATION NUMBER: US/09/248,796A

; CURRENT FILING DATE: 1999-02-12

; PRIOR APPLICATION NUMBER: US 60/074,725

; PRIOR FILING DATE: 1998-02-13

; PRIOR APPLICATION NUMBER: US 60/096,409

; PRIOR FILING DATE: 1998-08-13

; NUMBER OF SEQ ID NOS: 28208

; SEQ ID NO 21233

; LENGTH: 310

; TYPE: PRT

; ORGANISM: Candida albicans

US-09-248-796A-21233

Query Match 10.2%; Score 352; DB 2; Length 310;

Best Local Similarity 29.6%; Pred. No. 3.1e-30;

Matches 95; Conservative 35; Mismatches 97; Indels 94; Gaps 8;

QY	144	LITFSYSLTMIPFNRNVDOACSYLSQDGLGVADFYVS-----GKYD-----LPLRQMP	193
DB	3	LITFSYSLSMIPTFNAAIDNAVSKLDMEGIIATVDFGIQSSDTSMGRIINTVGGLVNRDIP	62
QY	194	WRRRPFWRIFIDINIDIGPERAVLEOKLERWEQNTQ-----GSIPYVPLRAPPYVWI	249
DB	63	WILRNFWRIFWADKVFLDSRRNRYLEYKFGTVKSLNSYNKALGKI-----PYIWI	114
QY	250	GRLPVSVGHALHEE---RVERPMPFPPT-----	273
DB	115	GDCKSKSHILERLNCLATESPYLATPTTPIANQLEDIPISKHEAALINLQNLPPSM	174
QY	274	-----FLYTOSWEDPEPDMVMEINPKDTVLTLTSGCNAL	309
DB	175	YYOKEYRWVYDEMNPVYBQFNQYIYFTWEDPRDHKLNFSTDTVLTAITSAGDNIL	234
QY	310	NL--LVQAGQVVSVDNCPAQSAALLEKKVAQQLQEFEDVWOLFGEVHPRIELEYKLL	367
DB	235	SVASLPTPKKHDAVDLNPQNHLLKLASFCLSQEQIWSMGFGEGKIENFNDLIDTL	294
QY	368	APPLSQTSNFWSKRLWYFQH 388	
DB	295	APHMSNA-----FQH 305	

```
RESULT 2
US-09-248-796A-26335
; Sequence 26335, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 26335
; LENGTH: 222
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-26335

Query Match 4.6%; Score 158; DB 2; Length 222;
Best Local Similarity 35.1%; Pred. No. 9.7e-09;
Matches 34; Conservative 20; Mismatches 19; Indels 24; Gaps 3;

Qy 51 LESFYGQAAAF-----AARLAERSNLIWDLGGGTGGENVDMADYI 92
Db 123 LESFYKQAAHIYONTREFLLKGRQCLRLAISHLPKKKDLIWDIGGGTGSNIEFMDEIS 182
Qy 93 DLAK-FKSIYVVDLCHSLCEVAKKAKA-----KGWK 123
Db 183 KISENFAYIVDLSPLCEVAKARFESHDLTPRDWK 219

RESULT 3
US-09-107-532A-7254
; Sequence 7254, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 7254:
```

```
SEQUENCE CHARACTERISTICS:
LENGTH: 212 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc_feature
LOCATION: (B) LOCATION 1...212
SEQUENCE DESCRIPTION: SEQ ID NO: 7254:
US-09-107-532A-7254

Query Match 3.5%; Score 121; DB 2; Length 212;
Best Local Similarity 30.8%; Pred. No. 0.00013;
Matches 33; Conservative 23; Mismatches 44; Indels 8; Gaps 3;

Qy 73 IWVDLGGGTGGENVDMADYIDLAKFKSIYVVDLCHSLCEVAKKAKAKGKNNVQVVEADA 132
Db 53 VLLDYGSGTGLVSLPLAE-----RFKELIADASETMLKWAEEKIQAADLKNVRTIHADA 107
Qy 133 -COFAPPEGTATLITFSYSLTMIPPHNVVDQACSYLSQDGLGVADF 179
Db 108 SVEF--PAVQANLILSLVLLHIPDPTENILTKLYEILAPGGQLIIVDF 153

RESULT 4
US-09-902-540-16360
; Sequence 16360, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 16360
; LENGTH: 329
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-09-902-540-16360

Query Match 3.4%; Score 118; DB 2; Length 329;
Best Local Similarity 21.8%; Pred. No. 0.00059;
Matches 81; Conservative 37; Mismatches 121; Indels 132; Gaps 16;

Qy 264 VERPPMFPPTFLYTQSWEDPEPDMVMEINPKDVTLTSTGGCNALMLVQ-GAGQVVSV 322
Db 6 VSTPPL---RLKEAVVREDAALELALVERTRARAVLTVASGGCTLLTLARRHPALELVGF 62
Qy 323 DCNPAQSALLELKKVAIQQL-----EFEDVHQLFG-----EGVHPRIEELYEKKLAPFL 371
Db 63 DFNPROLAHVREKAEGRLPLARYSDAEDAALNORGEFGLFRTLRRFIEEFVAP-- 120
Qy 372 SQTSHNFWSKRLWYFQGLYYQGMGKLCWVLQGLAVLGLGKTVKELANAP-TMBEQRR 430
Db 121 ---AHE-----LAFF-----APATTASQRR 138
Qy 431 LWDNSMLIHFKNGPKPLVWLFKVLSVLFNKAVLWFGGVPKQKVALIKADGIPENY 490
Db 139 -----EACARWFAS--PYWPVAFELALAAPLN- 164
Qy 491 IARTMDGVAENSHVRKQNY-----FYNCLTGKFLRDNCPYLR 529
Db 165 ---TMFGPAATQHAEPGSGYPGYQAVFERGLQREDAPRNPFLQHVLLGLRDLREDAPEYLR 221
```

QY 530 -BAFA-TLKGVDNLTSTNFMELKARTYTKVILMDHV-DWLDMPVANELAECLAK 586  
DB 222 AEGALATLVQGLPDVPRDLRF-----DVLSLSNIFDWSDALVAEWAGYLAR 270  
QY 587 QVAPGGIVWR 597  
DB 271 EAPGCAVLIR 281

## RESULT 5

US-09-583-110-4761  
; Sequence 4761, Application US/09583110  
; Patent No. 6699703  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al.  
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus  
; FILE REFERENCE: Pneumoniae for Diagnostics and Therapeutics  
; CURRENT APPLICATION NUMBER: US/09/583,110  
; CURRENT FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/107,433  
; PRIOR FILING DATE: 1998-06-30  
; PRIOR APPLICATION NUMBER: US 60/085,131  
; PRIOR FILING DATE: 1998-05-12  
; PRIOR APPLICATION NUMBER: US 60/051,553  
; PRIOR FILING DATE: 1997-07-02  
; NUMBER OF SEQ ID NOS: 5322  
; SEQ ID NO 4761  
; LENGTH: 195  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-583-110-4761

Query Match 3.2%; Score 111; DB 2; Length 195;  
Best Local Similarity 28.0%; Pred. No. 0.0014;  
Matches 40; Conservative 28; Mismatches 59; Indels 16; Gaps 6;

QY 46 DHAARLESFVGPO---AAFAARLAER-----SNLIWDLGGGTGENVDMADYIDLAK- 96  
DB 4 DFNHKAETDPSKNIPLANLVCQAVEKQIDILSDKVIDLDFGGGTG---LLA--LPLAKQ 57  
QY 97 FKSIVYVDLCHSLCEVAKKAKAGKWKVQVVEADACQFAPPGTATLTFTSYSLTWIPP 156  
DB 58 AKSVTLVDISEKMLEQARLKVEQQAIKNTQFLQDLPK-NPLEKEFDCLAVSRVLHMPD 116  
QY 157 FHNVIDOACSYLSQDGLGVADF 179  
DB 117 LDAALSFLFHQHLKEDGKLIADF 139

## RESULT 6

US-09-107-433-4622  
; Sequence 4622, Application US/09107433  
; Patent No. 6800744  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID  
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 5206  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: <Unknown>  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: <Unknown>  
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,433  
FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/ 085131  
FILING DATE: May 12, 1998  
APPLICATION NUMBER: 60/051553  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Ariniello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-011  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 4622:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 198 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Streptococcus pneumoniae  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: (B) LOCATION 1...198  
SEQUENCE DESCRIPTION: SEQ ID NO: 4622:  
US-09-107-433-4622

Query Match 3.2%; Score 111; DB 2; Length 198;  
Best Local Similarity 28.0%; Pred. No. 0.0015;  
Matches 40; Conservative 28; Mismatches 59; Indels 16; Gaps 6;  
QY 46 DHAARLESFVGPO---AAFAARLAER-----SNLIWDLGGGTGENVDMADYIDLAK- 96  
DB 7 DFNHKAETDPSKNIPLANLVCQAVEKQIDILSDKVIDLDFGGGTG---LLA--LPLAKQ 60  
QY 97 FKSIVYVDLCHSLCEVAKKAKAGKWKVQVVEADACQFAPPGTATLTFTSYSLTWIPP 156  
DB 61 AKSVTLVDISEKMLEQARLKVEQQAIKNTQFLQDLPK-NPLEKEFDCLAVSRVLHMPD 119  
QY 157 FHNVIDOACSYLSQDGLGVADF 179  
DB 120 LDAALSFLFHQHLKEDGKLIADF 142

## RESULT 7

US-09-252-991A-24766  
; Sequence 24766, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 24766  
; LENGTH: 706  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-24766

Query Match 3.1%; Score 108.5; DB 2; Length 706;  
Best Local Similarity 19.6%; Pred. No. 0.026;  
Matches 94; Conservative 54; Mismatches 170; Indels 161; Gaps 25;  
QY 42 KKGDDHAARLESFYGPQAAFAARLAERSN-----LIWVDLGGGTGENVDMADYI 92

```
Db 174 REGDELLQRCARGGPPIAAACAPLAERLQACQSVBERVALM--LAAQTGELLRLVGPRE 231
Qy 93 DLAKFSIYVDLCHSLCEVAKKAKAKAGWNV--QVVEADACQAPPBGT-----A 142
Db 232 HLVRSEYYPV-----EVERPLARPASWYELFPRESGDPTR-----HGTFFDVIRRLP 279
Qy 143 TLITFSYSLTWIPFHNV-----IQACSYL--SQDG----- 172
Db 280 QIAAMGFDVLYFFPIHPIGTHRKNNSLRAGADPGSPYAIGSDGGHEATHPELGDR 339
Qy 173 -----LVGV-----ADFYVSGKYDLP-LRQMP--WSRRFFWR-----SI 203
Db 340 EDFRLLIVA VREHGMELADFQICSPDHPWLEHPCW---FAWRPDGSLRYAENPPKY 396
Qy 204 FDTNDIDIGERRAYLEOKLERWEQNTQGSIPYV-----PWLRL-APYYWI-- 249
Db 397 EDIVNVDF-----YAEQALPSLWEALRDVVLGWEQGVTLFRVDNPHTKPLPFWEWLIA 450
Qy 250 ---GRLPVSGHALHEERVERPPM-----FPPTFLY-----TQSW 280
Db 451 EVRGRHPQV--IFLSEAFTRPAMMARLGKVGFSQSYTYFTWRNDKQELABYFAELNQPPW 508
Qy 281 ED---PEPDMEVMEINPKDTVTLTSGGCNALLVQAG-----QVVSVDNCPAQ 329
Db 509 RDCVRPNFFVTDPINPWFQSRGRGFLTRAALATMGSLGWMYSGFELCEAAALPGKE 568
Qy 330 ALLELKKVAIQOEFDVWQFEGVHPRIEELYEKK--LAPFLSQTSNFWKRLWYF 386
Db 569 EYLDSEKYLPRDYQAPGNIVAE--IARLNRIIRRENPAIQTLGFOAYNAMNDRIYF 625

RESULT 8
US-09-667-373-2
; Sequence 2, Application US/09667373
; Patent No. 6524840
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Nehls, Michael C.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: Enzyme-Like Proteins and Polynucleotides Encoding the Same
; FILE REFERENCE: LEX-0043-USA
; CURRENT APPLICATION NUMBER: US/09/667,373
; PRIOR FILING DATE: 2000-09-22
; PRIOR FILING DATE: 1999-09-24
; PRIOR FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 255
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-667-373-2

Query Match 3.1%; Score 107; DB 2; Length 255;
Best Local Similarity 22.9%; Pred. No. 0.0065;
Matches 53; Conservative 20; Mismatches 68; Indels 90; Gaps 11;

Qy 213 PERR-AYLEOKLERVWEQNTQGSIPYVWLRAPYYVWIGRLPSVGHALHEERVERPPMFP 271
Db 14 PERNCGYREVE---YWDQRYQGAADSAP-----YDWFQDFSSFRALL----- 52
Qy 272 PTFLYTQSWEDPEPDMEVMEINPKDTVTLTSGGCN-----ALNLLVOGAGQVVSVCNPA 327
Db 53 -----EP-----ELRPEDRILVL---GCGNSALSVELFLGGFPNVTSDYSSV 92
Qy 328 QSALLELKKVAIQOELEFE-----DWWOLFGEVGVH 356
Db 93 VVAAMQARYAHVPQLRWETMDVKLDPPSPASFDVLEKGTLDALLAGERDPMTVSSEGVH 152
Qy 357 PRIEELYE--KKLAP---FLSQTS-----HNFWSKRLWYFQHGILYYQG 394
Db 153 TVDQVLSEVSRVLVPGGRFISMSTAAPHPRTRHYAQAYYGWSLRHATYGS 203

RESULT 9
US-10-305-413-2
; Sequence 2, Application US/10305413
; Patent No. 6835564
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Nehls, Michael C.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: Enzyme-Like Proteins and Polynucleotides Encoding the Same
; FILE REFERENCE: LEX-0043-USA
; CURRENT APPLICATION NUMBER: US/10/305,413
; CURRENT FILING DATE: 2002-11-25
; PRIOR FILING DATE: US/09/667,373
; PRIOR FILING DATE: 2000-09-22
; PRIOR FILING DATE: 1999-09-24
; PRIOR FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 255
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-305-413-2

Query Match 3.1%; Score 107; DB 2; Length 255;
Best Local Similarity 22.9%; Pred. No. 0.0065;
Matches 53; Conservative 20; Mismatches 68; Indels 90; Gaps 11;

Qy 213 PERR-AYLEOKLERVWEQNTQGSIPYVWLRAPYYVWIGRLPSVGHALHEERVERPPMFP 271
Db 14 PERNCGYREVE---YWDQRYQGAADSAP-----YDWFQDFSSFRALL----- 52
Qy 272 PTFLYTQSWEDPEPDMEVMEINPKDTVTLTSGGCN-----ALNLLVOGAGQVVSVCNPA 327
Db 53 -----EP-----ELRPEDRILVL---GCGNSALSVELFLGGFPNVTSDYSSV 92
Qy 328 QSALLELKKVAIQOELEFE-----DWWOLFGEVGVH 356
Db 93 VVAAMQARYAHVPQLRWETMDVKLDPPSPASFDVLEKGTLDALLAGERDPMTVSSEGVH 152
Qy 357 PRIEELYE--KKLAP---FLSQTS-----HNFWSKRLWYFQHGILYYQG 394
Db 153 TVDQVLSEVSRVLVPGGRFISMSTAAPHPRTRHYAQAYYGWSLRHATYGS 203

RESULT 10
US-09-489-039A-13088
; Sequence 13088, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLBSSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
```

```
Db 93 VVAAMQARYAHVPQLRWETMDVKLDPPSPASFDVLEKGTLDALLAGERDPMTVSSEGVH 152
Qy 357 PRIEELYE--KKLAP---FLSQTS-----HNFWSKRLWYFQHGILYYQG 394
Db 153 TVDQVLSEVSRVLVPGGRFISMSTAAPHPRTRHYAQAYYGWSLRHATYGS 203

RESULT 9
US-10-305-413-2
; Sequence 2, Application US/10305413
; Patent No. 6835564
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Nehls, Michael C.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: Enzyme-Like Proteins and Polynucleotides Encoding the Same
; FILE REFERENCE: LEX-0043-USA
; CURRENT APPLICATION NUMBER: US/10/305,413
; CURRENT FILING DATE: 2002-11-25
; PRIOR FILING DATE: US/09/667,373
; PRIOR FILING DATE: 2000-09-22
; PRIOR FILING DATE: 1999-09-24
; PRIOR FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 255
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-305-413-2

Query Match 3.1%; Score 107; DB 2; Length 255;
Best Local Similarity 22.9%; Pred. No. 0.0065;
Matches 53; Conservative 20; Mismatches 68; Indels 90; Gaps 11;

Qy 213 PERR-AYLEOKLERVWEQNTQGSIPYVWLRAPYYVWIGRLPSVGHALHEERVERPPMFP 271
Db 14 PERNCGYREVE---YWDQRYQGAADSAP-----YDWFQDFSSFRALL----- 52
Qy 272 PTFLYTQSWEDPEPDMEVMEINPKDTVTLTSGGCN-----ALNLLVOGAGQVVSVCNPA 327
Db 53 -----EP-----ELRPEDRILVL---GCGNSALSVELFLGGFPNVTSDYSSV 92
Qy 328 QSALLELKKVAIQOELEFE-----DWWOLFGEVGVH 356
Db 93 VVAAMQARYAHVPQLRWETMDVKLDPPSPASFDVLEKGTLDALLAGERDPMTVSSEGVH 152
Qy 357 PRIEELYE--KKLAP---FLSQTS-----HNFWSKRLWYFQHGILYYQG 394
Db 153 TVDQVLSEVSRVLVPGGRFISMSTAAPHPRTRHYAQAYYGWSLRHATYGS 203

RESULT 10
US-09-489-039A-13088
; Sequence 13088, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLBSSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
```

```
; SEQ ID NO 13088
; LENGTH: 677
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-13088

Query Match      3.0%; Score 104.5; DB 2; Length 677;
Best Local Similarity 21.4%; Pred. No. 0.068;
Matches 63; Conservative 28; Mismatches 91; Indels 111; Gaps 15;

QY 235 IPYVPLRAPYVWIGRLPSVGHALH-----EERVERPMPFP-PTLYTQSWEDPE 284
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 193 IEYSPGVGDYWIWALQSLGTTLTGTFINFEVILIKRAGPMFMKVPV-----SWASLC 248
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 285 PDMEVMEINPKDVTLTSTGGCNALNLLVQAGQVSDVCNPAQSALLEKKAIVAOOLEF 344
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 249 ANILLIATSP---ILTVT-----IALLTLDRYLGTHTFFT 279
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 345 ED-----VWQLFGGVHPRIEELYEKKLAPF--LSQTSHPNFWSKRLWYFQGLY 391
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 280 NDGMGNMMYINLIW-AWG---HP---EVYILVLPVGFVSEIATFSKRLP----- 325
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 392 YQGMGKLCVLOCIAV-----LGLGKTVKRLANAPTEBQRRLWDSNMLIHF 440
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 326 --GYSLWATVCIIVLSFIWVLLHFFTWGAGANVNAFFGTTM-----IIA 370
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 441 VKNGKPLVWLFVKFVSLVFNKAVLW-----FGG-----GVPGKQYAL 479
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 371 IPTGVKIFNWLFTMYQGRIVFNSAMMTIGFIVTFSVGGMTGVLAVPGADFVL 424
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 11
US-09-107-532A-3716
; Sequence 3716, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3716:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1070 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
```

```
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...1070
; SEQUENCE DESCRIPTION: SEQ ID NO: 3716:
US-09-107-532A-3716

Query Match      3.0%; Score 103.5; DB 2; Length 1070;
Best Local Similarity 22.2%; Pred. No. 0.2;
Matches 80; Conservative 54; Mismatches 113; Indels 113; Gaps 18;

QY 260 HEERVERPMPFPPTLYTQSWEDPEPEMEVMEINPKDVTLTSTGGCNALNLLVQGA--- 316
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 491 HOPKIE-----VTETGSLDIRFDVTGIERQEIDHVLQ-----SLLRNDAFYT 533
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 317 ---GOVSVDCNPAQ--SALLEKKVAIQ-----OLE-----PEDVW 348
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 534 LENGQVLSFDEEFQOTSQVLQOLRESIRTEGTHVPRKNQGLIIQNOLKSNATSESF 593
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 349 Q-LFGGVHPRIEELYEKKLAPLSQTSNHF-----WSKRLWYFQHLGYQGMGKLCW 401
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 594 QTWQDLIHP---ERYQAQLPKGLNATMRDYQKQGRFLKWLGHYQFG----- 638
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 402 VLQCLAVVLGLGKTVKRLANAPTEBQRRLWDSNMLIHFVKNQPKPLVW-----LFVK 454
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 639 --GILADEGLGKTLQTIAPLLSEKERKFSALIV-----APASLIYNWQAEVRKFAP 690
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 455 FVSLVLFNKAFLWFGGVGKQYALIKADGIPIENTARTMDGVAENSHVRKQNYFYINC 514
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 691 SLISQVIN-----GNKKEREELLAKDTRIVTSYASLRQD--LANYQSKIDYL---I 738
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 515 LTGKFLRDNCP-----YLREAA-----FATLKSGVDNL-----TVSTNFEELKAR 558
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 739 LDEAQVKNSSKTAQALRELAVPQRFALSGTFIENNLBELSLFATIMPFPTTKFR 798
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 12
US-09-711-164-332
; Sequence 332, Application US/09711164
; Patent No. 6589738
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE
; FILE REFERENCE: ELITRA.008A
; CURRENT APPLICATION NUMBER: US/09/711,164
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/164415
; PRIOR FILING DATE: 1999-11-9
; NUMBER OF SEQ ID NOS: 469
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 332
; LENGTH: 663
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-711-164-332

Query Match      2.9%; Score 101.5; DB 2; Length 663;
Best Local Similarity 22.2%; Pred. No. 0.14;
Matches 63; Conservative 35; Mismatches 95; Indels 91; Gaps 15;

QY 235 IPYVPLRAPYVWIGRLPSVGHALHEERVERPMPFPPTLYTQSWEDPEPEMEVMEINP 294
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 179 IEYSPGVGDYWIWLSQLSGIGTTLTGIN-----FFVTLKKRA-----PGMTMKM-P 226
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 295 KDTVLTLTSGGCNANLNLVQAGQVSDVCNPAQSALLEKKVAIQQLRFED----- 346
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 227 VFTWASLCA-----NVLIIASFPIITVTV-----ALLTLDRYLGTHTFTNDGMGNM 275
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 347 ---VMQLFGGVHPRIEELYEKKLAPF--LSQTSHPNFWSKRLWYFQHLGYQGMGKLCW 401
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

THERET



Db	276	INLIW-AWG---HP---EVILLIPVGVFSEIAATFSRKLFP-----GYTSLVW 318	QY	198	FWRSIFDIDNIDIGERRAYLEQKLERVMEQNTQGSIPYVWMLRAPYYVWIGRLFSVGH 257
QY	402	VLOCLAVW-----LGLCKTVKRLANAPTMEQRRLDSNMLLIHFVKNGPKPLVW 450	Db	195	YGGDAHFDD-----ETWSSKG-----YNLFVAHEFGH 256
Db	319	ATVCITVLSPFVWLHFFTMGAGANVAFGIITM-----IIAIPGVKIFNW 366	QY	258	ALHEERVERPP--MPDPTFLYQSMEDPDPMEVMEI-----NPKDTVLTLTSG 304
QY	451	LFVKFVSLVLFNKAVLW-----FGG-----GVPGKQVAL 479	Db	227	SLGLDHSKDPGALMF--PIYTYTGKSHFMLPDDDDVOGISOGLYGFGEDEDPNPKH----- 277
Db	367	LFTMYOGRIVFHSAMLWTIGFIVTFSVGGMTGVLVAVPGADFVL 410	QY	305	GCNALNLLVQAGQOVVSDCNPAQS--ALLELKKVAIOOLEPED--VWOLFEGGVHP--- 357
RESULT 13			Db	278	-----PKTPDKDPSLSLDAITSURG---ETMIFKDRFFWRL-----HPQOV 316
US-08-994-689C-1			QY	358	-----RIBELYEKKLAPFLSQTSHNFWSKRLWYFQHGLYIOGGMKLCW 401
; Sequence 1, Application US/08994689C			Db	317	DAELFLTCKSPWELPNRIDAAYE-----HPSHDL-----IFIFR-----GRKFW 355
; Patent No. 6613958			QY	402	VLOCLAVW-----LGLGKTVKRLANAPTMEQRRLWDSNMLLIHFVKNGPKPLVWLF 452
; GENERAL INFORMATION:			Db	356	ALNGYDILEGYPKKISELGLPKVEVKISAA-----VHFEDTG----- 392
; APPLICANT: Neuhold, Lisa			QY	453	VKFVSLVLFNKAVLW---EGGGVPGKQYALIKADGIPENYIARTWDGVAENSHVRKQNY 509
; APPLICANT: Killar, Loran			Db	393	-----KTLFSGNQWRYDDTNHMDKYPRLIEEDFP---GIGDKVDAYVE-----KNGY 440
; TITLE OF INVENTION: TRANSGENIC ANIMAL MODEL FOR			QY	510	FYV 512
; TITLE OF INVENTION: DEGENERATIVE DISEASES OF CARTILAGE			Db	441	IYF 443
; NUMBER OF SEQUENCES: 21			RESULT 14		
; CORRESPONDENCE ADDRESS:			US-09-134-078-24		
; ADDRESSEE: Darby & Darby PC			; Sequence 24, Application US/09134078		
; STREET: 805 Third Avenue			; Patent No. 636844		
; CITY: New York			; GENERAL INFORMATION:		
; STATE: NY			; APPLICANT: Bylina, Edward J.		
; COUNTRY: USA			; TITLE OF INVENTION: GLYCOSIDASE ENZYMES		
; ZIP: 10022			; NUMBER OF SEQUENCES: 72		
; COMPUTER READABLE FORM:			; CORRESPONDENCE ADDRESS:		
; MEDIUM TYPE: Diskette			; ADDRESSEE: Gray Cary Ware & Freidenrich LLP		
; COMPUTER: IBM Compatible			; STREET: 4365 Executive Drive, Suite 1600		
; OPERATING SYSTEM: DOS			; CITY: San Diego		
; SOFTWARE: FastSeq for Windows Version 2.0			; STATE: CA		
; CURRENT APPLICATION DATA:			; COUNTRY: USA		
; FILING DATE: 1997-12-19			; ZIP: 92121		
; CLASSIFICATION: 800			; COMPUTER READABLE FORM:		
; PRIOR APPLICATION DATA:			; MEDIUM TYPE: Diskette		
; APPLICATION NUMBER:			; COMPUTER: IBM Compatible		
; FILING DATE:			; OPERATING SYSTEM: Windows95		
; ATTORNEY/AGENT INFORMATION:			; SOFTWARE: FastSeq for Windows Version 2.0		
; NAME: Green, Reza			; CURRENT APPLICATION DATA:		
; REGISTRATION NUMBER: 38,475			; APPLICATION NUMBER: US/09/134,078		
; REFERENCE/DOCKET NUMBER: 0630/0D532			; FILING DATE: 13-AUG-1998		
; TELEPHONE: 212-527-7700			; CLASSIFICATION: 435		
; TELEFAX: 212-753-6237			; PRIOR APPLICATION DATA:		
; TELEX: 236687			; APPLICATION NUMBER: 08/949,026		
; INFORMATION FOR SEQ ID NO: 1:			; FILING DATE: 10-OCT-1997		
; SEQUENCE CHARACTERISTICS:			; APPLICATION NUMBER: 60/056,916		
; LENGTH: 471 amino acids			; FILING DATE: 06-DEC-1996		
; TYPE: amino acid			; ATTORNEY/AGENT INFORMATION:		
; STRANDEDNESS: single			; NAME: Haile, Lisa A.		
; TOPOLOGY: linear			; REGISTRATION NUMBER: 38,347		
; MOLECULE TYPE: protein			; REFERENCE/DOCKET NUMBER: 09010/024002		
US-08-994-689C-1			; TELECOMMUNICATION INFORMATION:		
Query Match 2.9%; Score 99.5; DB 2; Length 471;			; TELEPHONE: 858/677-1456		
Best Local Similarity 18.4%; Pred. No. 0.13;			; TELEFAX: 858/677-1456		
Matches 100; Conservative 61; Mismatches 159; Indels 223; Gaps 27;			; INFORMATION FOR SEQ ID NO: 24:		
QY	51	LESFYGP-----QAAFAARLAERNLIWDLGGGTGEN-VDM-----ADYDLA 95	QY	258	ALHEERVERPP--MPDPTFLYQSMEDPDPMEVMEI-----NPKDTVLTLTSG 304
Db	43	LRSYTHPTNLAGILKENAASMTLRMQSFGLVETGKLDNDTLSDVMKKPRGVVDVG 102	Db	227	SLGLDHSKDPGALMF--PIYTYTGKSHFMLPDDDDVOGISOGLYGFGEDEDPNPKH----- 277
QY	96	KFKSI-----YVVDLCHSICEVAKKAKAKGKNVQVVEADACQAPP 138	QY	305	GCNALNLLVQAGQOVVSDCNPAQS--ALLELKKVAIOOLEPED--VWOLFEGGVHP--- 357
Db	103	EYNVFPRLKWSKNLYRINYVTPDMTHSEVEKAFKA-FKWSVDYTPIN----- 152	Db	278	-----PKTPDKDPSLSLDAITSURG---ETMIFKDRFFWRL-----HPQOV 316
QY	139	EGTATLITFSYSLTMIPFFHNVIDOACSYLSQDGLGVGVADFY-VSGKYDLPRLQMPWSRR 197	QY	358	-----RIBELYEKKLAPFLSQTSHNFWSKRLWYFQHGLYIOGGMKLCW 401
Db	153	-----FTRLHDGIADIMISFGIKHGDYFPDGFSGLLAHAFPPGPN 194	Db	317	DAELFLTCKSPWELPNRIDAAYE-----HPSHDL-----IFIFR-----GRKFW 355



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2006, 01:59:26 ; Search time 15 Seconds  
(without alignments)  
1202.467 Million cell updates/sec

Title: US-10-620-914-45  
Perfect score: 3463  
Sequence: 1 MGSGRDPASVTKNKNFSL.....RVNMYSSFYMARCKGAKXDN 648

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 161667 seqs, 27834885 residues

Total number of hits satisfying chosen parameters: 161667

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_AA\_New.\*

- 1: /cgn2\_6/prodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/prodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/prodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/prodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 5: /cgn2\_6/prodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/prodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 7: /cgn2\_6/prodata/1/pubpaa/US11\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/prodata/1/pubpaa/US60\_NEW\_PUB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	106	3.1	472	7	US-11-169-041-156
2	102	2.9	448	7	US-11-124-367A-282
3	96	2.8	326	6	US-10-467-657-7224
4	95.5	2.8	2890	7	US-11-115-639-31
5	95.5	2.8	2890	7	US-11-115-639-32
6	95.5	2.8	2890	7	US-11-115-639-33
7	94	2.7	658	7	US-11-096-568A-2582
8	94	2.7	680	7	US-11-096-568A-2581
9	94	2.7	680	7	US-11-096-568A-2580
10	93.5	2.7	1197	6	US-10-055-877-8
11	93.5	2.7	1247	6	US-10-055-877-10
12	93	2.7	417	7	US-11-098-686-11330
13	92.5	2.7	1590	6	US-10-055-877-146
14	92	2.7	316	7	US-11-096-568A-1265
15	91.5	2.6	241	7	US-11-098-686-10262
16	91	2.6	749	7	US-11-072-512-2622
17	89	2.6	537	7	US-11-109-156-28
18	89	2.6	577	6	US-10-718-264-3
19	89	2.6	577	6	US-10-718-264-3
20	88.5	2.6	329	7	US-11-234-786-376
21	87.5	2.5	834	7	US-11-087-099-10005
22	87	2.5	265	6	US-10-467-657-6590
23	87	2.5	265	6	US-10-467-657-7654
24	87	2.5	565	7	US-11-072-512-2180
25	86.5	2.5	395	6	US-10-793-626-664

ALIGNMENTS

RESULT 1

US-11-169-041-156  
; Sequence 156, Application US/11169041  
; Publication No. US20060019284A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: IDENTIFICATION OF POLYNUCLEOTIDES FOR PREDICTING ACTIVITY OF  
; TITLE OF INVENTION: COMPOUNDS THAT INTERACT WITH AND/OR MODULATE PROTEIN TYROSINE  
; TITLE OF INVENTION: KINASES AND/OR PROTEIN TYROSINE KINASE PATHWAYS IN LUNG CANCER  
; TITLE OF INVENTION: CELLS  
; FILE REFERENCE: 10001 NP  
; CURRENT APPLICATION NUMBER: US/11/169,041  
; CURRENT FILING DATE: 2005-06-28  
; PRIOR APPLICATION NUMBER: 60/584,405  
; PRIOR FILING DATE: 2004-06-30  
; NUMBER OF SEQ ID NOS: 527  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 156  
; LENGTH: 472  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-169-041-156

Query Match 3.1%; Score 106; DB 7; Length 472;  
Best Local Similarity 20.4%; Pred. No. 0.13;  
Matches 90; Conservative 54; Mismatches 155; Indels 142; Gaps 19;

QY	12	YTKQNFLEKLSNMKDDLTVLRHMWFGSKGDDHAA-----RLRSFYGPQAA--AFAA 64
DB	39	FYRTFQNFREFK-ATMNCNLLAVLKHLL-----KGNEAALECLRKABELTQOEADQAEIR 92
QY	65	RLAERSNLWDLGGGTGENVDMDADYID--LAKFKSIYVVDICSLCEVAKKAKAGW 122
DB	93	SLVTWGNVAVVYHMGRLSDVQIYVDVKVHCKEFPYIESPELDCBEGWTRLCGGN 152
QY	123	KN--VOVVEADACQAP--PEGTATLITFSYSLTMTIPPHNVIDQACSYLSQDGLGVAD 178
DB	153	QNERAKVCEKALEKKPKNPEFTSGLAIASYRLDNWPPSQNAID-----196
QY	179	FYVSGKYDPLQMPWRSRFFWRSIFDIDNIDIGPER---RAYLEOKLRVWVFQNGSI 235
DB	197	-----PLRQA-----IRLNPONQYLVLLALKLHKHREE-----225
QY	236	PYVPMLRABYYVYVWIGRLPSVGHALHEERVERPP---MFPPTFLYTSQWEDPFDMEV--289
DB	226	-----GEEGEKEKLVVEALEKAPGVTDLRSAAKFFYRKDEPDKALIELLK 271
QY	290	--MEINPKDTVLTLTSGGCNALLNVQAGQVVSVDNCP--AOSALLELKVAIQOL---342

Db 272 KALEYIPNNAYLHCQIGCCYRAKVF-----QVNNLRNGMYGKRKLELIGHAVAHKKA 326  
QY 343 -----BFDVWQFGBGVHPRIEYKELAPLSQTSF----- 376  
Db 327 DEANDNLFRVCSILASHALADQYDAEYVF-----QKEFSKELTPVAKQLLHRYG 378  
QY 377 NFV-----SKRLWYFQHGL 390  
Db 379 NFQLYQMKCEDKAIHHFIEGV 399

RESULT 2  
US-11-124-367A-282  
; Sequence 282, Application US/11124367A  
; Publication No. US20060024700A1  
; GENERAL INFORMATION:  
; APPLICANT: Michele Cargill  
; APPLICANT: Hongjin Huang  
; TITLE OF INVENTION: Genetic Polymorphisms Associated with  
; TITLE OF INVENTION: Fibrosis Methods of Detection and Uses Thereof  
; FILE REFERENCE: CLO01519.ORD  
; CURRENT APPLICATION NUMBER: US/11/124, 367A  
; CURRENT FILING DATE: 2005-05-09  
; PRIOR FILING DATE: 2005-05-09  
; PRIOR FILING DATE: 2004-05-07  
; PRIOR APPLICATION NUMBER: US 60/582,609  
; PRIOR FILING DATE: 2004-06-25  
; PRIOR APPLICATION NUMBER: US 60/599,554  
; PRIOR FILING DATE: 2004-08-09  
; NUMBER OF SEQ ID NOS: 34460  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 282  
; LENGTH: 448  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-124-367A-282

Query Match 2.9%; Score 102; DB 7; Length 448;  
Best Local Similarity 20.8%; Pred. No. 0.29;  
Matches 90; Conservative 54; Mismatches 163; Indels 126; Gaps 19;

QY 12 YTKKNSLEKLSMKDDTLVLRHWMFGSKGDDHAA-----RLSFYGPQAA--AFAA 64  
Db 15 FYTFEQNRFFK-ATMCNLLAYLKLH-----KQNEAALECKAKAEELIQEHADQAEIR 68  
QY 65 RLAEARNLIWVGGGGTGENVDMADYID--LAKFSIYVVDLCHSLCEVAKKKAKAGW 122  
Db 69 SLVTWGNVAVVYHMGRLSDVQIYVDKVVHCEKFSFSPYRIESPELDCBEGWRLKCGGN 128  
QY 123 KN--VQVVEADACQAP--PEGTATLTFSYSLTMIPPFHNVIDQACSYLSQDGLGVAD 178  
Db 129 QNERAKVCFBKALEKPKNPEFTSGLAISYRLDNWPPSQNAID----- 172  
QY 179 FYVSGKYDLPURQWPSRRFRFWSIFDIDNIDIGPER---RAYLEOKLVRVWEQNTQGS 235  
Db 173 -----PLRQA-----IRLNPQYKLVLLALKHMRSE----- 201  
QY 236 PYVPWLRAPYVVMIGRLPSVGHALHEERVRPP-----MFPPTLYTQSWEDPDPMEV-- 289  
Db 202 -----GEBEGEKLVEALEKAPGVTVLRSAAKFFYRRKDEPKAIELLK 247  
QY 290 --MEINPKDVTILTSGGCNALNLLVQAGQVVSVCNP--AQSALELEAKKVAIQOLEFE 345  
Db 248 KALEYIPNNAYLHCQIGCCYRAKVF-----QVNNLRNGMYGKRKLELIGHAVAHKKA 302  
QY 346 D-----VWQLFG-----EGVHPRIEELYEKKLAPFLSQTSF-----NFW----- 379  
Db 303 DEANDNLFRVCSILASHALADQYEAEEYVFQKEFSKELTPVAKQLLHRYGNFQYQMK 362  
QY 380 --SKRLWYFQHGL 390  
Db 363 CEDKAIHHFIEGV 375

RESULT 3  
US-10-467-657-7224  
; Sequence 7224, Application US/10467657  
; Publication No. US20050260581A1  
; GENERAL INFORMATION:  
; APPLICANT: CHIRON SPA  
; APPLICANT: FONTANA Maria Rita  
; APPLICANT: PIZZA Mariagrazia  
; APPLICANT: MASNANI Vega  
; APPLICANT: MONACI Elisabetta  
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS  
; FILE REFERENCE:  
; CURRENT APPLICATION NUMBER: US/10/467,657  
; CURRENT FILING DATE: 2003-08-11  
; PRIOR APPLICATION NUMBER: GB-0103424.8  
; PRIOR FILING DATE: 2001-02-12  
; NUMBER OF SEQ ID NOS: 9218  
; SOFTWARE: SeqWin99, version 1.04  
; SEQ ID NO 7224  
; LENGTH: 326  
; TYPE: PRT  
; ORGANISM: Neisseria gonorrhoeae  
US-10-467-657-7224

Query Match 2.8%; Score 96; DB 6; Length 326;  
Best Local Similarity 25.9%; Pred. No. 0.64;  
Matches 30; Conservative 21; Mismatches 39; Indels 26; Gaps 6;

QY 120 KGKKNVQVVEADACQAPP---EGTATLTFSYSLTMIPPF-----HNV--IDOACS--- 166  
Db 214 EGWLEHEVRLTVCEYLKPLADGIDTLVLGCTHFFLLKPLIGREAHNVAVDSAITTAE 273  
QY 167 ----YLSQDLGVAD-----FYSGKYDLPURQWPSRRFRFWSIFDIDNIDIG 212  
Db 274 ETARVLAQEGLLDTGNNDYRPFVS---DIPLKPRFTIGERFLGRTMEQIEMVSLG 326

RESULT 4  
US-11-115-639-31  
; Sequence 31, Application US/11115639  
; Publication No. US20050282242A1  
; GENERAL INFORMATION:  
; APPLICANT: Rothstein, David  
; APPLICANT: Murphy, Christopher  
; APPLICANT: MacNeil, Ian  
; TITLE OF INVENTION: SCREENING ASSAYS FOR ANTIMICROBIAL  
; TITLE OF INVENTION: AGENTS  
; FILE REFERENCE: 50150/075003  
; CURRENT APPLICATION NUMBER: US/11/115,639  
; CURRENT FILING DATE: 2005-04-27  
; PRIOR APPLICATION NUMBER: 60/566,858  
; PRIOR FILING DATE: 2004-04-30  
; PRIOR APPLICATION NUMBER: 60/565,679  
; PRIOR FILING DATE: 2004-04-27  
; NUMBER OF SEQ ID NOS: 55  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 31  
; LENGTH: 2890  
; TYPE: PRT  
; ORGANISM: H. pylori  
US-11-115-639-31

Query Match 2.8%; Score 95.5; DB 7; Length 2890;  
Best Local Similarity 18.9%; Pred. No. 17;  
Matches 101; Conservative 82; Mismatches 150; Indels 201; Gaps 29;

QY 214 ERRAY-----LEOKLERYWE---ONTQGSIPYVPWLRAPYVVMIGRLPSVGHALH 260  
Db 1775 EERGYATTLKQAKRMIEQKSNVEWECLOETEG-----YFVLLNRAFT-----LH 1819  
QY 261 EERVERPPMPPPTFLYTQSWEDPEPDMVMEINP----- 294

Db 1820 KOSIQ-----AFHPKLI-----DKAIQLHPLVCSAFNADFDGDMVHVPLSQA 1866  
Qy 295 -KDTVLTGSGCNALNLVQAGQVSVDCNPAQSALLEKKVAIQLEFEDVWOLFGE 353  
Db 1867 AECVKMLSS-----WNILLPASGKAVAI---PSQDVLGLYLSLEKSGVGEHKLFS 1918  
Qy 354 -----GVHPRIEELYE-----KKLAPLSOTSNFWSKRLWYFQ 387  
Db 1919 VNEIITAITDKELDIHAKIRVLDQGNIIATISAGRMIIKSILPDFIPT--DLWNRPMKKD 1976  
Qy 388 HG-----LYYQGMGKLCWVLOCL-----AVVLGLG-----KTVKRLANAPTME 426  
Db 1977 IGVLDVYVHKVGIGITATFLDNLKTGLGFRYATKAGISISMEDIIPTPKQKQKVEKAKVE 2036  
Qy 427 EQR--RLMDSNMLIHFWKNGPKPLVWLFV-----KFVSLVL-----FNKAVLWFGGGVP 473  
Db 2037 VKKIQQYQGLLTDOERYNKIIDTWTENVDKSKEMMTAIAQDKGFSIYMMADSGAR 2096  
Qy 474 GK-----QYALIKADGPIPIENYIARTMDGVAENSHVRKQNYFYNCCLTG--KFLR 521  
Db 2097 GSAAIQIRQLSAMRGLMTKPDGSIITETPI---ISNFKEGNLVLE---YFNSTHGARKGLA 2149  
Qy 522 DNCPTVLRGAAPATLKSQVVDNLTVSTNF-----FMEELKAR 558  
Db 2150 D---TALKTNAGYLTRKLIID---VSNQVSVDDCGTGHGIEITDIAVGSSELIEPLEER 2203  
Qy 559 TYTKVILMDHVDLWDMPVANEL---AECL-----AKQVAPGGIVWRSASLSPP 604  
Db 2204 IFRVLELVID-----PITNEILLYADTLIDEGAKKVVVEAGI---KSITIRTP 2250

RESULT 5

US-11-115-639-32  
; Sequence 32, Application US/11115639  
; Publication No. US2005028242A1  
; GENERAL INFORMATION:  
; APPLICANT: Rothstein, David  
; APPLICANT: Murphy, Christopher  
; APPLICANT: MacNeill, Ian  
; TITLE OF INVENTION: SCREENING ASSAYS FOR ANTIMICROBIAL  
; FILE REFERENCE: 50150/075003  
; CURRENT APPLICATION NUMBER: US/11/115,639  
; PRIOR FILING DATE: 2005-04-27  
; PRIOR APPLICATION NUMBER: 60/566,858  
; PRIOR FILING DATE: 2004-04-30  
; PRIOR APPLICATION NUMBER: 60/565,679  
; PRIOR FILING DATE: 2004-04-27  
; NUMBER OF SEQ ID NOS: 55  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 32  
; LENGTH: 2890  
; TYPE: PRT  
; ORGANISM: H. pylori  
US-11-115-639-32

Query Match 2.8%; Score 95.5; DB 7; Length 2890;  
Best Local Similarity 18.9%; Pred. No. 17;  
Matches 101; Conservative 82; Mismatches 150; Indels 201; Gaps 29;  
Qy 214 ERROR-----LEOKLERWE---QNTQGSIPYVWLRAPYVWITGRPLPSVGHALH 260  
Db 1775 EERGATTLKQAKRMIEQKSNWECLQITEG-----YFVLLNRAPT---LH 1819  
Qy 261 EERVERPPMPPTFLYTSQWEDPEPDMVEINP----- 294  
Db 1820 KOSIQ-----AFHPKLI-----DKAIQLHPLVCSAFNADFDGDMVHVPLSQA 1866  
Qy 295 -KDTVLTGSGCNALNLVQAGQVSVDCNPAQSALLEKKVAIQLEFEDVWOLFGE 353  
Db 1867 AECVKMLSS-----WNILLPASGKAVAI---PSQDVLGLYLSLEKSGVGEHKLFS 1918

Qy 354 -----GVHPRIEELYE-----KKLAPLSOTSNFWSKRLWYFQ 387  
Db 1919 VNEIITAITDKELDIHAKIRVLDQGNIIATISAGRMIIKSILPDFIPT--DLWNRPMKKD 1976  
Qy 388 HG-----LYYQGMGKLCWVLOCL-----AVVLGLG-----KTVKRLANAPTME 426  
Db 1977 IGVLDVYVHKVGIGITATFLDNLKTGLGFRYATKAGISISMEDIIPTPKQKQKVEKAKVE 2036  
Qy 427 EQR--RLMDSNMLIHFWKNGPKPLVWLFV-----KFVSLVL-----FNKAVLWFGGGVP 473  
Db 2037 VKKIQQYQGLLTDOERYNKIIDTWTENVDKSKEMMTAIAQDKGFSIYMMADSGAR 2096  
Qy 474 GK-----QYALIKADGPIPIENYIARTMDGVAENSHVRKQNYFYNCCLTG--KFLR 521  
Db 2097 GSAAIQIRQLSAMRGLMTKPDGSIITETPI---ISNFKEGNLVLE---YFNSTHGARKGLA 2149  
Qy 522 DNCPTVLRGAAPATLKSQVVDNLTVSTNF-----FMEELKAR 558  
Db 2150 D---TALKTNAGYLTRKLIID---VSNQVSVDDCGTGHGIEITDIAVGSSELIEPLEER 2203  
Qy 559 TYTKVILMDHVDLWDMPVANEL---AECL-----AKQVAPGGIVWRSASLSPP 604  
Db 2204 IFRVLELVID-----PITNEILLYADTLIDEGAKKVVVEAGI---KSITIRTP 2250

RESULT 6

US-11-115-639-33  
; Sequence 33, Application US/11115639  
; Publication No. US2005028242A1  
; GENERAL INFORMATION:  
; APPLICANT: Rothstein, David  
; APPLICANT: Murphy, Christopher  
; APPLICANT: MacNeill, Ian  
; TITLE OF INVENTION: SCREENING ASSAYS FOR ANTIMICROBIAL  
; FILE REFERENCE: 50150/075003  
; CURRENT APPLICATION NUMBER: US/11/115,639  
; PRIOR FILING DATE: 2005-04-27  
; PRIOR APPLICATION NUMBER: 60/566,858  
; PRIOR FILING DATE: 2004-04-30  
; PRIOR APPLICATION NUMBER: 60/565,679  
; PRIOR FILING DATE: 2004-04-27  
; NUMBER OF SEQ ID NOS: 55  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 33  
; LENGTH: 2890  
; TYPE: PRT  
; ORGANISM: H. pylori  
US-11-115-639-33

Query Match 2.8%; Score 95.5; DB 7; Length 2890;  
Best Local Similarity 18.9%; Pred. No. 17;  
Matches 101; Conservative 82; Mismatches 150; Indels 201; Gaps 29;  
Qy 214 ERROR-----LEOKLERWE---QNTQGSIPYVWLRAPYVWITGRPLPSVGHALH 260  
Db 1775 EERGATTLKQAKRMIEQKSNWECLQITEG-----YFVLLNRAPT---LH 1819  
Qy 261 EERVERPPMPPTFLYTSQWEDPEPDMVEINP----- 294  
Db 1820 KOSIQ-----AFHPKLI-----DKAIQLHPLVCSAFNADFDGDMVHVPLSQA 1866  
Qy 295 -KDTVLTGSGCNALNLVQAGQVSVDCNPAQSALLEKKVAIQLEFEDVWOLFGE 353  
Db 1867 AECVKMLSS-----WNILLPASGKAVAI---PSQDVLGLYLSLEKSGVGEHKLFS 1918  
Qy 354 -----GVHPRIEELYE-----KKLAPLSOTSNFWSKRLWYFQ 387  
Db 1919 VNEIITAITDKELDIHAKIRVLDQGNIIATISAGRMIIKSILPDFIPT--DLWNRPMKKD 1976  
Qy 388 HG-----LYYQGMGKLCWVLOCL-----AVVLGLG-----KTVKRLANAPTME 426  
Db 1977 IGVLDVYVHKVGIGITATFLDNLKTGLGFRYATKAGISISMEDIIPTPKQKQKVEKAKVE 2036

QY 427 EQR--RLWDSNMLIHFKVNGPKPLVWLV-----KPVSLVL-----FNKAVLWFGGVP 473  
Db 2037 VKKIQOYDGLLTDQERYNKIITDTEVNDKMSKEMTAIAQDKEGFSNIYMWADSGAR 2096  
QY 474 GK-----QYALIKADGIPYIARTMDGVAENSHVRKONTFYNYNCLTG--KFLR 521  
Db 2097 GSAAQIRQLSAMRGLMTKPDGSIETPI---ISNFKELNVLE---YFNSTHGARKGLA 2149  
QY 522 DNCPTVIREAFATLKSQVVDNUTVSTNF-----PMESLKAR 558  
Db 2150 D---TALKTANAGYLRKLID---VSONVKVSDDCGTHGEIBITDIAVGSSELIEPLEER 2203  
QY 559 TYTKVILMDHVDLMDPVANEL---AECL-----AKQVAPGGIVIVRSASLSPP 604  
Db 2204 IFGRVLEEDVID---PITNEILLYADTLIDBEGAKKVVEAGI---KSIITIRTP 2250

RESULT 7  
US-11-096-568A-2582  
; Sequence 2582, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096, 568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 2582  
; LENGTH: 658  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(658)  
; OTHER INFORMATION: Ceres Seq. ID no. 15176030  
US-11-096-568A-2582

Query Match 2.7%; Score 94; DB 7; Length 658;  
Best Local Similarity 19.6%; Pred. No. 2.7;  
Matches 115; Conservative 69; Mismatches 182; Indels 220; Gaps 30;

QY 20 EKLKLSMKDDLTVLRHWMFGSKKGD-----DHAARLESFYGPQAAFAARLAER 69  
Db 56 EKLQLAGVSQRCVVL-----GSKAEDKHLQLVHTKDHNVLVKSISTKQDYRNRNASQ 110  
QY 70 SNLIWDLGGGTGENDVMADYIDLAKFKSIYVVDLCHSLCEVAKKAKAKGKNVQVVE 129  
Db 111 LNSIY--LNGSSEAYLAAG-----SVVKLAEKVAG--E 142  
QY 130 ADACQAPPEGTATLITFSYSLTMIPPFNHVIDOACSYLSQDGLGVADVFSYKGYDLPL 189  
Db 143 LD--CGFA-----IVRPPGHAADEAMGFCFLFNNVAVAAASFLNERPDGV 187  
QY 190 RQ---MPW-----SRPFWR-----IFDIDNIDIGPERRAYLEOKLERVWEQNTQGI 235  
Db 188 KKLIVDMVHNGTQKMFWDKPRVLFFSVHRHEVGGFYAGDDGDDYNNVVGSGPGEFN 247  
QY 236 PYVPWLRAPIYVWIGRLPSVGHALHEERVERPPM--FPPTFLYTSQSWEDPEPD----- 286  
Db 248 INVPDQ-----GRCDADYLAADHILIPVAREFNPDIIVLSAGFADAAINDPLGGCC 300  
QY 287 -----MEVNEINPKDVTLTSTGGCNALNLVQAGQVVSVDNCNPAQSALELKK 336  
Db 301 VTPYGYVMLKKLMEFAQGIKIVLAL--EGGYN--LDSIAKSSSLACVQV-----LLEDKQ 350  
QY 337 V--AIQOLEFEDVWQFGEVHPRI-----EEL-----YEKKLAPFL-----SOTS 375  
Db 351 IQGPPAYPESTWRVI--QAVRKCLCTYWPSLADELSWKLINQKTPPTPIILISSDSETE 409  
QY 376 HN-----FWSKRL-----WYFQHG-----L 390

Db 410 DNAQGLLDQMSKLSIENPOGTLLENHQVEPASTSWRADLAKVDVWYASFGSNMWKPRFLC 469  
QY 391 YYQGGMGKLCWLVLOCLAVVLGLKTVKRLANAPTMEEQRELWDSNMLIHFKVNGPKPLVW 450  
Db 470 YIQGGQ-----VDGLKKVC-----VGSMD-----KSPPKETVW 497  
QY 451 LFWKFSVSLVFNK--AVLWFGGVPKGQYALIKADGIPYIART 494  
Db 498 --ETFPRLFFRGRESSVGVGV-----AFTNPLANLIDQT 532

RESULT 8  
US-11-096-568A-2581  
; Sequence 2581, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096, 568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 2581  
; LENGTH: 660  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(660)  
; OTHER INFORMATION: Ceres Seq. ID no. 15176029  
US-11-096-568A-2581

Query Match 2.7%; Score 94; DB 7; Length 660;  
Best Local Similarity 19.6%; Pred. No. 2.7;  
Matches 115; Conservative 69; Mismatches 182; Indels 220; Gaps 30;

QY 20 EKLKLSMKDDLTVLRHWMFGSKKGD-----DHAARLESFYGPQAAFAARLAER 69  
Db 58 EKLQLAGVSQRCVVL-----GSKAEDKHLQLVHTKDHNVLVKSISTKQDYRNRNASQ 112  
QY 70 SNLIWDLGGGTGENDVMADYIDLAKFKSIYVVDLCHSLCEVAKKAKAKGKNVQVVE 129  
Db 113 LNSIY--LNGSSEAYLAAG-----SVVKLAEKVAG--E 144  
QY 130 ADACQAPPEGTATLITFSYSLTMIPPFNHVIDOACSYLSQDGLGVADVFSYKGYDLPL 189  
Db 145 LD--CGFA-----IVRPPGHAADEAMGFCFLFNNVAVAAASFLNERPDGV 189  
QY 190 RQ---MPW-----SRPFWR-----IFDIDNIDIGPERRAYLEOKLERVWEQNTQGI 235  
Db 190 KKLIVDMVHNGTQKMFWDKPRVLFFSVHRHEVGGFYAGDDGDDYNNVVGSGPGEFN 249  
QY 236 PYVPWLRAPIYVWIGRLPSVGHALHEERVERPPM--FPPTFLYTSQSWEDPEPD----- 286  
Db 250 INVPDQ-----GRCDADYLAADHILIPVAREFNPDIIVLSAGFADAAINDPLGGCC 302  
QY 287 -----MEVNEINPKDVTLTSTGGCNALNLVQAGQVVSVDNCNPAQSALELKK 336  
Db 303 VTPYGYVMLKKLMEFAQGIKIVLAL--EGGYN--LDSIAKSSSLACVQV-----LLEDKQ 352  
QY 337 V--AIQOLEFEDVWQFGEVHPRI-----EEL-----YEKKLAPFL-----SOTS 375  
Db 353 IQGPPAYPESTWRVI--QAVRKCLCTYWPSLADELSWKLINQKTPPTPIILISSDSETE 411  
QY 376 HN-----FWSKRL-----WYFQHG-----L 390  
Db 412 DNAQGLLDQMSKLSIENPOGTLLENHQVEPASTSWRADLAKVDVWYASFGSNMWKPRFLC 471  
QY 391 YYQGGMGKLCWLVLOCLAVVLGLKTVKRLANAPTMEEQRELWDSNMLIHFKVNGPKPLVW 450  
Db 472 YIQGGQ-----VDGLKKVC-----VGSMD-----KSPPKETVW 499



QY 451 LFVKFVSLVLFNK--AVLWFGGVPKGQYALIKADGIPNIYART 494  
Db 500 --ETFFHRLFFGREGSVGNGGV-----AFTNPLANLIDQT 534

RESULT 9  
US-11-096-568A-2580  
; Sequence 2580, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; FILE OF INVENTION: Therby  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096, 568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 2580  
; LENGTH: 680  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; NAME/KEY: misc feature  
; LOCATION: (1)..(680)  
; OTHER INFORMATION: Ceres Seq. ID no. 15176028  
US-11-096-568A-2580

Query Match 2.7%; Score 94; DB 7; Length 680;  
Best Local Similarity 19.6%; Pred. No. 2.8;  
Matches 115; Conservative 69; Mismatches 182; Indels 220; Gaps 30;

QY 20 EKLKLSMKDDTLVLRHMFSGKGD-----DHAARLESFYGPQAAFAARLAER 69  
Db 78 EKLQAGVSRCVVL-----GSSKAEDKHLQLVHTKDHVNLVKSISTKQDYERNIASQ 132

QY 70 SNLIWDLGGCTGENTVDMDADYIDLAKFSIYVDLCHSLCEFAKKAQKAKGKWNQVVE 129  
Db 133 LNSIY--LNGSGSEAYLAAG-----SVVLAERVAEG-----E 164

QY 130 ADACQAPPEGTATLTFYSVLTMIPEHNVIDQACSVLSQDGLGVADPVYSGKYDLPL 189  
Db 165 LD--CGFA-----IVRPPGHAEADENMGCLFNNAVVAASFLMLNERPD LGV 209

QY 190 RQ--MPW-----SRRPFWR-----IFDIDNIDIGPERRAYLEOKLERVWEQNTQSSI 235  
Db 210 KKLILVDMDVHNGCTQKMPKDPRLVFTSVHRHEYGFPAGDDGDNVNVGSGPGEFN 269

QY 236 PYPVWLRAPYYWIGRLPSVGHALHEERVERPPM--PPFTFLYTQSWDEPPD----- 286  
Db 270 INVPMQ-----GRCGDADYLAADWHILIPVAREFNPDIPLSAGFDDAAINDPLGGCC 322

QY 287 -----MEVWEINPKDTVLTLTSGGNCNALLVQAGQVVSVDNCPAQSALLLELK 336  
Db 323 VTPYGSVMKLKMEFAQGIKVLAL--EGGYN--LDSIAKSLACVQV-----LLEDKQ 372

QY 337 V--AIQOLEPEDVQWLFGEVHPRI-----EEL-----YKKLAPFL-----SOTS 375  
Db 373 IQGPPAYPESTWRVI-QAVRKLCTYWPSLADELSWKLINGKTPTPILISSDSETE 431

QY 376 HN-----FWSKRL-----WYFQHG-----L 390  
Db 432 DNAQGLLDQMSKLSIENPQGLLENHQVEPASTSRADLAKVDVWVASFGSNMKNKPRFLC 491

QY 391 YYQGGMGKLCWVLOCLAVVLGKTVKRLANAPTMEEQRLDWSNMLIHFKNGKPLW 450  
Db 492 YIQGGQ-----VDGLKKVC-----VGSMD-----KSPPKETW 519

QY 451 LFVKFVSLVLFNK--AVLWFGGVPKGQYALIKADGIPNIYART 494  
Db 520 --ETFFHRLFFGREGSVGNGGV-----AFTNPLANLIDQT 554

RESULT 10  
US-10-055-877-8  
; Sequence 8, Application US/10055877  
; Publication No. US20050288241A1  
; GENERAL INFORMATION:  
; APPLICANT: DeCristofaro, Marc  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Miller, Charles  
; APPLICANT: Tchernev, Velizar  
; APPLICANT: Zhong, Mei  
; APPLICANT: Anderson, David  
; APPLICANT: Ballinger, Robert  
; APPLICANT: Gerlach, Valerie  
; APPLICANT: Spytek, Kimberly  
; APPLICANT: Ratelli, Luca  
; APPLICANT: Kekuda, Ramesh  
; APPLICANT: Guo, Xiaojia  
; APPLICANT: Zerhusen, Bryan  
; APPLICANT: Andrew, David  
; APPLICANT: Mezes, Peter  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Burgess, Cahterine  
; APPLICANT: Eissen, Andrew  
; APPLICANT: Wolenc, Adam  
; APPLICANT: Baumgartner, Jason  
; APPLICANT: Shimkets, Richard  
; APPLICANT: Gusev, Vladimir  
; APPLICANT: Vernet, Corine  
; APPLICANT: Taupier Jr., Raymond  
; APPLICANT: Pena, Carol  
; APPLICANT: Shenoy, Suresh  
; APPLICANT: Li, Li  
; APPLICANT: Casman, Stacie  
; APPLICANT: Boldog, Ference  
; TITLE OF INVENTION: Novel Polypeptides and Nucleic Acids Encoded Thereby  
; FILE REFERENCE: 21402-251  
; CURRENT APPLICATION NUMBER: US/10/055,877  
; CURRENT FILING DATE: 2002-01-22  
; PRIOR APPLICATION NUMBER: 60/262,892  
; PRIOR FILING DATE: 2001-01-19  
; PRIOR APPLICATION NUMBER: 60/263,598  
; PRIOR FILING DATE: 2001-01-23  
; PRIOR APPLICATION NUMBER: 60/263,799  
; PRIOR FILING DATE: 2001-01-24  
; PRIOR APPLICATION NUMBER: 60/264,117  
; PRIOR FILING DATE: 2001-01-25  
; PRIOR APPLICATION NUMBER: 60/264,139  
; PRIOR FILING DATE: 2001-01-25  
; PRIOR APPLICATION NUMBER: 60/264,478  
; PRIOR FILING DATE: 2001-01-26  
; PRIOR APPLICATION NUMBER: 60/263,351  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: 60/272,870  
; PRIOR FILING DATE: 2001-03-02  
; PRIOR APPLICATION NUMBER: 60/275,990  
; PRIOR FILING DATE: 2001-03-14  
; PRIOR APPLICATION NUMBER: 60/275,927  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 512  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 8  
; LENGTH: 1197  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-055-877-8

Query Match 2.7%; Score 93.5; DB 6; Length 1197;  
Best Local Similarity 21.3%; Pred. No. 7.1;  
Matches 51; Conservative 28; Mismatches 81; Indels 79; Gaps 9;

QY 227 WEQNTQGSIPYVWLRAPYYWIGRLPSVGHALHEERVERPPMFPPT----- 273

Db 342 WERLASSTAPYIPELRGP-----MDTSNFDVDDDTLNHPGTLPPPSHGAFSGHHLPFV 394  
QY 274 -FLYTOSWEDPDPMEVMEINPKDVTTLTSGGCNALLVQAGQVSVDCNPAQSALL 332  
Db 395 GFTYTSAAALERKLCQLEQ-----KLPAGGSPQLRKEVAALRBOLEQAHSRRLEQEA 448  
QY 333 ELKKVAIQ-----LEFEDVMQLFGEVHPRIELEYEKKLAPFLSQTSHNFWSKRLW 384  
Db 449 EKOSQALQOELAMLREBLEQESKQRL--EGERRETSNWEAQLADILS-----494  
QY 385 YFQHGLYYQGGMGKLCWV-----LQCLAVVLGLGKTVKRLANAPTMEEQ---RRL 431  
Db 495 -----WVNDKVSRCYLOALAT--KMAEELSIRNVGTDHGWKARRL 535

## RESULT 11

US-10-055-877-10

; Sequence 10, Application US/10055877

; Publication No. US20050288241A1

; GENERAL INFORMATION:

; APPLICANT: DeCristofaro, Marc

; APPLICANT: Padigar, Muralidhara

; APPLICANT: Miller, Charles

; APPLICANT: Tchernev, Velizar

; APPLICANT: Zhong, Mei

; APPLICANT: Anderson, David

; APPLICANT: Ballinger, Robert

; APPLICANT: Gerlach, Valerie

; APPLICANT: Spytek, Kimberly

; APPLICANT: Ratelli, Luca

; APPLICANT: Kekuda, Ramesh

; APPLICANT: Guo, Xiaojia

; APPLICANT: Zerhusen, Bryan

; APPLICANT: Andrew, David

; APPLICANT: Mezes, Peter

; APPLICANT: Patturajan, Meera

; APPLICANT: Burgess, Catherine

; APPLICANT: Eisen, Andrew

; APPLICANT: Wolenc, Adam

; APPLICANT: Baumgartner, Jason

; APPLICANT: Shimkets, Richard

; APPLICANT: Gusev, Vladimir

; APPLICANT: Vernet, Corine

; APPLICANT: Taupier Jr., Raymond

; APPLICANT: Pena, Carol

; APPLICANT: Shenoy, Suresh

; APPLICANT: Li, Li

; APPLICANT: Casman, Stacie

; APPLICANT: Boldog, Ference

; TITLE OF INVENTION: Novel Polypeptides and Nucleic Acids Encoded Thereby

; FILE REFERENCE: 21402-251

; CURRENT APPLICATION NUMBER: US/10/055,877

; CURRENT FILING DATE: 2002-01-22

; PRIOR APPLICATION NUMBER: 60/262,892

; PRIOR FILING DATE: 2001-01-19

; PRIOR APPLICATION NUMBER: 60/263,598

; PRIOR FILING DATE: 2001-01-23

; PRIOR APPLICATION NUMBER: 60/263,799

; PRIOR FILING DATE: 2001-01-24

; PRIOR APPLICATION NUMBER: 60/264,117

; PRIOR FILING DATE: 2001-01-25

; PRIOR APPLICATION NUMBER: 60/264,139

; PRIOR FILING DATE: 2001-01-25

; PRIOR APPLICATION NUMBER: 60/264,478

; PRIOR FILING DATE: 2001-01-26

; PRIOR APPLICATION NUMBER: 60/263,351

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: 60/272,870

; PRIOR FILING DATE: 2001-03-02

; PRIOR APPLICATION NUMBER: 60/275,990

; PRIOR FILING DATE: 2001-03-14

; PRIOR APPLICATION NUMBER: 60/275,927

; PRIOR FILING DATE: 2001-03-14

; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 512  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 10  
; LENGTH: 1247  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-055-877-10

Query Match 2.7%; Score 93.5; DB 6; Length 1247;

Best Local Similarity 21.3%; Pred. No. 7.5; Indels 79; Gaps 9;

Matches 51; Conservative 28; Mismatches 81

QY 227 WEONTGSGIPYVFWLRAPIYVVMIGRLPSVGHALHEERVERPPMPPT-----273

Db 342 WERLASSTAPYIPELRGP-----MDTSNFDVDDDTLNHPGTLPPPSHGAFSGHHLPFV 394

QY 274 -FLYTOSWEDPDPMEVMEINPKDVTTLTSGGCNALLVQAGQVSVDCNPAQSALL 332

Db 395 GFTYTSAAALERKLCQLEQ-----KLPAGGSPQLRKEVAALRBOLEQAHSRRLEQEA 448

QY 333 ELKKVAIQ-----LEFEDVMQLFGEVHPRIELEYEKKLAPFLSQTSHNFWSKRLW 384

Db 449 EKOSQALQOELAMLREBLEQESKQRL--EGERRETSNWEAQLADILS-----494

QY 385 YFQHGLYYQGGMGKLCWV-----LQCLAVVLGLGKTVKRLANAPTMEEQ---RRL 431

Db 495 -----WVNDKVSRCYLOALAT--KMAEELSIRNVGTDHGWKARRL 535

## RESULT 12

US-11-098-686-11330

; Sequence 11330, Application US/11098686

; Publication No. US20060024696A1

; GENERAL INFORMATION:

; APPLICANT: Kapur, Vivek and Gebhart, Connie J.

; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES

; FILE REFERENCE: 09531-128001

; CURRENT APPLICATION NUMBER: US/11/098,686

; CURRENT FILING DATE: 2005-04-04

; PRIOR APPLICATION NUMBER: PCT/US03/31318

; PRIOR FILING DATE: 2003-10-01

; PRIOR APPLICATION NUMBER: US 60/416,395

; PRIOR FILING DATE: 2002-10-04

; NUMBER OF SEQ ID NOS: 11433

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 11330

; LENGTH: 417

; TYPE: PRT

; ORGANISM: Lawsonia intracellularis

US-11-098-686-11330

Query Match 2.7%; Score 93; DB 7; Length 417;

Best Local Similarity 21.8%; Pred. No. 1.7;

Matches 58; Conservative 43; Mismatches 87; Indels 78; Gaps 16;

QY 56 GPOAAAFARLAERSNLI-----WVDLGGGTGENVDMADYIDLAK-----EKSIVVVDL 105

Db 16 GMTAAAPAAKGGKCLITKNGTTIGSGT---IDVIG-YLPGKVPVSNPFEAVTFPK 71

QY 106 CHS---LCEVAKKA-----KAKGWKNVQVVEADACQAPPEGT---ATL 144

Db 72 RHPYQLIGELTVRNALKNFSELVGEENVYKTSQKNYVVT-----AGTIKPTWM 123

QY 145 ITFSYSLTMTIPPHNVIDQACSVLQDGLVGADVFVSGKYDPLROMPNSRRFFWTSIF 204

Db 124 VTESMDMSIIPDIKNI---CIL---GIEGMKDFP---ELIKGLS-SKKYFANKTF 170

QY 205 DIDNIDIGPER-----AYLEQKLERVW-----EONTQGSIPVVMRAPYVWIG 250

Db 171 TQAKLHFPESERNISTNVQAYLETENGLKWLVGKIQQGNQSDLIILPILGTTF-----226



```
Qy 211 IGPERRAYLEOKLERVWEQNTQGSIPVVPWLRAPYYVWIGRLPSVGHALHEERVPMPF 270
Db 202 ---ORRAAEQEQRLMRQ-----ARMNVDIEKLMS-----NMSIERIDSF 240
Qy 271 PPTFLYTQSWEDPEPDMEVMEINP-----KDTVLTLTSGGCNALNLVLVQAGAGQVVSVD--- 323
Db 241 DAVERYL-----LRENE-NPGKGSADSI-----GACQKLKVDLQAAAFNVLPKDLQ 285
Qy 324 ---CNPAQSALLE 333
Db 286 LLLSNPKRAVLLQ 298
```

## RESULT 15

```
US-11-098-686-10262
; Sequence 10262, Application US/11098686
; Publication No. US20060024696A1
; GENERAL INFORMATION:
; APPLICANT: Kapur, Vivek and Gebhart, Connie J.
; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES
; FILE OF INVENTION: FROM LAWSONIA INTRACELLULARIS AND METHODS OF USING
; FILE REFERENCE: 09531-128001
; CURRENT APPLICATION NUMBER: US/11/098,686
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: PCT/US03/31318
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: US 60/416,395
; PRIOR FILING DATE: 2002-10-04
; NUMBER OF SEQ ID NOS: 11433
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 10262
; LENGTH: 241
; TYPE: PRT
; ORGANISM: Lawsonia intracellularis
US-11-098-686-10262
```

Query Match 2.6%; Score 91.5; DB 7; Length 241;

Best Local Similarity 20.5%; Pred. No. 1.1; Matches 33; Conservative 29; Mismatches 70; Indels 29; Gaps 5;

```
Qy 79 GGTGENVDMMADYIDLA-----KPKSIYV--VDLCHSLCEVAKKAKGKXQVQVVEAD 131
Db 53 GKTGRILDLAAGTLDVAIALHNRVEQLHISAVDLCFSMLYQGMKCLHRANITRWPTAD 112
Qy 132 ACQFAPPEGATLITFSYSLTMTPPFNNVIDQACSYLSQDGLGVGVADPFVSGKYDLP 191
Db 113 AKLLPDPDSCIDGVTLAGIRNIVPRTKAFSEIARVLLPGRMAILEF-GTGKQRI---- 167
Qy 192 MPWSRRFFWRSIFD-----IDNIDIGPERRAYLEQKL 223
Db 168 -----WMGLNYLYTKILPFIKUSSDPSAYLYLKQSI 200
```

Search completed: March 14, 2006, 02:01:14

Job time : 17 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: March 14, 2006, 04:22:40 ; Search time 317 Seconds  
(without alignments)  
3633.628 Million cell updates/sec

Title: US-10-620-914-45

Perfect score: 3463

Sequence: 1 MGSGRDRPASVTKNFKSLE.....RVNYSSTFYARRKGAADN 648

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-O=/abss/ABSSWEB.spool/US10620914/runat\_13032006\_102046\_6830/app.query.fasta\_1  
-DB=Issued Patents NA QFMT=fastap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-HOST=abss908 -USBR=US10620914 @CGN 1 1 193 @runat\_13032006\_102046\_6830 -NCPUS=6  
-ICPU=3 -NO WMAP -NEG SCORES=0 -WAIT -DEPBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:\*  
1: /cgn2\_6/ptodata/1/ina/1 COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5 COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/H COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/PCTUS COMB.seq:\*  
7: /cgn2\_6/ptodata/1/ina/PP COMB.seq:\*  
8: /cgn2\_6/ptodata/1/ina/RE COMB.seq:\*  
9: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	Match	Length	DB	ID	Description
1	352	10.2	933	3	US-09-248-796A-7130		Sequence 7130, Ap
2	158	4.6	666	3	US-09-248-796A-12232		Sequence 12232, A
c	3	138.5	4.0	8832	3	US-09-902-540-984	Sequence 984, App
	4	126.5	3.7	990	3	US-09-902-540-9203	Sequence 9203, Ap
	5	121	3.5	639	3	US-09-107-532A-3600	Sequence 3600, Ap
	6	118.5	3.4	1155	3	US-09-439-313-373	Sequence 373, App
	7	118.5	3.4	1155	3	US-09-352-616A-373	Sequence 373, App
	8	118.5	3.4	1155	3	US-09-289-198-301	Sequence 301, App
	9	118.5	3.4	1155	3	US-09-636-215-373	Sequence 373, App

RESULT 1

US-09-248-796A-7130  
; Sequence 7130, Application US/09248796A

; Patent No. 6747137

; GENERAL INFORMATION:

; APPLICANT: Keith Weinstock et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS

; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.132

; CURRENT APPLICATION NUMBER: US/09/248,796A

; CURRENT FILING DATE: 1999-02-12

; PRIOR APPLICATION NUMBER: US 60/074,725

; PRIOR FILING DATE: 1998-02-13

; PRIOR APPLICATION NUMBER: US 60/096,409

; PRIOR FILING DATE: 1998-08-13

; NUMBER OF SEQ ID NOS: 28208

; SEQ ID NO 7130

; LENGTH: 933

; TYPE: DNA

; ORGANISM: Candida albicans

US-09-248-796A-7130

Alignment Scores:

Pred. No.: 6.02e-33  
Score: 352.00  
Percent Similarity: 40.5%  
Best Local Similarity: 29.6%  
Query Match: 10.2%  
DB: 3  
Length: 933  
Matches: 95  
Conservative: 35  
Mismatches: 97  
Indels: 94  
Gaps: 8

US-10-620-914-45 (1-648) x US-09-248-796A-7130 (1-933)

QY

144 LeuileThrPheSerTyrSerLeuThrMetIleProPheHisanValIleAspGln 163

7	Db	 TTGATTACTTTTTCGTATTCATTTGTCGATGATCCCAACTTTTCAATGTGCTATCATGAAT	66
164	Qy	AlaCysSerTyrLeuSerGlnAspGlyLeuValGlyValAlaAspPheTyrValSer	182
67	Db	GCTGTTCTTAATTAGATATGAAGGTATTTATGCCACTGTAGATTTTGGTATTCAAAGC	126
183	Qy	-----GlyLysTyrAsp-----LeuProLeuArgGlnMetPro	193
127	Db	AGTGACACCTCAATGGGTCTATTAAATACTGTGTGGTGGTTTGAACAGGACATCTCT	186
194	Qy	TrpSerArgArgPhePheTyrArgSerIlePheAspIleAspAsnIleAspIleGlyPro	213
187	Db	TGGATATTACGTAAATTTTGGAGAAATTTGGTTTGAAAGCTGATAAAGTGTTTGGATCT	246
214	Qy	GluArgArgAlaTyrLeuGluGlnLysLeuGluArgValTrpGluGlnAsnTrpGln	232
247	Db	TCAAGNAGAAACTATTTGGAATATAAATTTGGTACCGTCAAACTCTTGAATTCATACAAC	306
233	Qy	-----GlySerIleProTyrValProTrpLeuArgAlaProTyrValTrpIle	249
307	Db	AAGGCTTTGGGTAAATC-----CCCTATTATATTGGATTT	342
250	Qy	GlyArgLeuProSerValGlyHisAlaLeuHisGluGlu-----ArgValGlu	265
343	Db	GTTTGTGATAAATCAAATCACACCACTTTTAGAAAGATTGCAATTTGTAGCCACTGAA	402
266	Qy	ArgProProMetPheProProThr-----	273
403	Db	TCCCTTACCTTGCTCCAACTACAACTCCAACTCGCTAATCACTTTGAAGATATTCCAAT	462
273	Qy	-----	273
463	Db	TCTAAGTGTATGAAGTGTCTTTAATCACTTGCAGAAAAAATTTACCTTACCCTCATG	522
273	Qy	-----	273
523	Db	TACTATCAAAAGGAATATTGGAGAGTCTACTATGATGAAATGAATCATTTGATGAACAA	582
274	Qy	-----PheLeuTyrThrGlnSerTrpGluAspProGluProAspMetGluVal	289
583	Db	TTTAAAAACCAATACATTTATGCTTTCTCTGGGAAGATCTCTCGTGAAGATCATAACTT	642
290	Qy	MetGluIleAsnProLysAspThrValLeuThrLeuThrSerGlyGlyCysAsnAlaLeu	309
643	Db	TGGAATTTTACCAGTGATGATACTGTTTGGCTATTACTTCAGCTGGTGATAATAATTTTG	702
310	Qy	AsnLeu-----LeuValGlnGlyAlaGlyGlnValValSerValAspCysAsnProAla	327
703	Db	AGTTATGTAGTTTACCACACCACCAAAAAAGATTTCAATGCTGTGATCTTAAATCCATGT	762
328	Qy	GlnSerAlaLeuLeuGluLeuLysLysValAlaIleGlnGlnLeuGluPheGluAspVal	347
763	Db	CAAAACCATTTATTAGAATTGAAATTTGGCTAGTTTTTAGATGCTTTCTCAAGAACAAT	822
348	Qy	TrpGlnLeuPheGlyGluGlyValHisProArgIleGluGluLeuTyrGlnLysLysLeu	367
823	Db	TGGTCAATGTTTGGTGAAGGTAAAAATTCAGAAATTTCAATGATCTTTTGAATGATCTTTG	882
368	Qy	AlaProPheLeuSerGlnThrSerHisAsnPheTrpSerLysArgLeuTrpTyrPheGln	387
893	Db	GGCCGCACATGCTCTTAATGCC-----TTCCAA	912
388	His	388	
913	CAC	915	

## RESULT 2

US-09-248-796A-12232  
; Sequence 12232, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstock et al

```

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDATE
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 12232
; LENGTH: 666
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-12232

Alignment Scores:
Pred. No.: 2,74e-09 Length: 666
Score: 158.00 Matches: 34
Percent Similarity: 55.7% Conservative: 20
Best Local Similarity: 35.1% Mismatches: 19
Query Match: 4.6% Indels: 24
DB: 3 Gaps: 3

US-10-620-914-45 (1-648) x US-09-248-796A-12232 (1-666)

Qy 51 LeuGluSerPheTyGlyProGlnAlaAlaAlaPhe----- 62
Db 367 TTAGAGTCATCTCAAGAACCAAGCTCATATTTATGATACACTAGAGAAATTTTGTG 426
Qy 63 -----AlaAlaArgLeuAlaGluArgSerAsnLeu 72
Db 427 AAGGTAGACAAGAATGCTTAGATTAGCTATTTCCCATTTTACCCAAAAGAAAGATCTT 486
Qy 73 IleTrrValAspLeuGlyGlyThrGlyGluAsnValAspMetMetAlaAspTyrIle 92
Db 487 ATTTGGATAGATATTTGGTGGTGGAACTGGTTCCCAATATTGAATTCATGGATGAAATTAGT 546
Qy 93 AspLeuAlaLys---PheLysSerIleTyrValValAspLeuCysHisSerLeuCysGlu 111
Db 547 AAAATATCTGNAAACTTTAAAGCTGTTATTTGGTTGATCTTTCCCATCTTTTGTGTGAA 606
Qy 112 ValAlaLysLysAlaLysAla-----LysGlyTrrLys 123
Db 607 GTTGCTAAGCAAGATTTGAATCCCATGATTTGACCAACCGGACTGGAAA 657

RESULT 3
US-09-902-540-984/c
; Sequence 984, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 984
; LENGTH: 8832
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-984

Alignment Scores:
Pred. No.: 6,27e-05 Length: 8832
Score: 138.50 Matches: 97
Percent Similarity: 37.7% Conservative: 50
Best Local Similarity: 24.9% Mismatches: 138

```





Db 403 TCACACGCGCGAGGC-----CTGTGC----- 425  
Qy 410 euGlyLeuGlyLysThrValLysArgLeuAlaAsnAlaProThrMetGluGluGlnArgA 430  
Db 425 ----- 425  
Qy 430 rgLeuTrpAspSerAsnMetLeuLeuHisPheValLysAsnGlyProLysProLeuValT 450  
Db 426 -----CCGCTGGTTGCTTGGCCT-AC TGCCCGCTGGCCTTCGAGCT 467  
Qy 450 rpLeuPheValLysPheValSerLeuValLeuPheAsnLysAlaValLeuTrpPheGly- 469  
Db 468 GGCCCTGGCGCGCG-CTGCTGAACACCATGTTTGGTCCGCGCGCAGCAGCATGCCG 526  
Qy 470 -----GlyGlyValProGlyLysGlnTrpAlaLeuLeuLysAlaAspGlyLeuProileg 488  
Db 527 AGCCCGGCTCGTATCTCTGCTATTTCAGCGCGTCTTCGAAACG----GGGCTCAGCGCG 583  
Qy 488 luAsnTrpIleAlaArgThrMetAspGlyValAlaGluAsnSerHisValArgLysGlnA 508  
Db 584 AGGAT-----GCGCTCGAA 598  
Qy 508 snTrpPheTrpTyrrAsnCysLeuThrGlyLysPheLeuArgAspAsnCysProThrTrpL 528  
Db 599 ATCCCTCTCCAGCAGCTGCTGCGGAGTACCTGCTGAGGATGCCCGGAGTATC 658  
Qy 528 euArg-----GluAlaAlaPheAla---ThrLeuLysSerGlyValValAspAsnLeuThrV 546  
Db 659 TCCGGGCTGAAGGCCCTGCGCTGACGCTGCTGCGGAGTGTCCCGAGTGTGCCG 718  
Qy 546 alSerThrAsnPhePheMetGluGluLeuLysAlaArgThrTrpLysValIleLeuM 566  
Db 719 GCCTGGACCGCTTC-----GACGTCACTCTCGC 745  
Qy 566 etAspHisVal---AspTrpLeuAspMetProValAlaAsnGluLeuAlaGluCysLeuA 585  
Db 746 TCTCCACATCTTCGACTGTGCGAAGACGCGCTGTGTGCGGAGTGTGCGGGGTGCTCG 805  
Qy 585 lalysGlnValAlaProGlyLysLeuValIleTrpArg-Ser----- 598  
Db 806 CAGGTGAGCGCGCGCGTGTGCGCGCTCTCTCATCCCGCAGCTCAACACCGCGCGGACC 865  
Qy 599 -----AlaSerLeuSerProTrpAlaGluLeuIleGlnLysAla 612  
Db 866 TTGCGCGCTCTTCACGCGCGCTTCGAGTTCGATGACGCGCTCGCGG 913  
  
RESULT 5  
US-10-620-914-45 (1-648) x US-09-107-532A-3600 (1-639)  
; Sequence 3600, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD/ROM ISO9660  
; COMPUTER: PC  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,532A  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Ariniello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 3600:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 639 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...639  
SEQUENCE DESCRIPTION: SEQ ID NO: 3600:  
US-09-107-532A-3600  
  
Alignment Scores:  
Pred. No.: 9,28e-05 Length: 639  
Score: 121.00 Matches: 33  
Percent Similarity: 51.9% Conservative: 23  
Best Local Similarity: 30.6% Mismatches: 44  
Query Match: 3.5% Indels: 8  
DB: 3 Gaps: 3  
  
US-10-620-914-45 (1-648) x US-09-107-532A-3600 (1-639)  
Qy 73 IleTrpValAspLeuGlyGlyThrGlyGluAsnValAspMetMetAlaAspTyrIle 92  
Db 157 GTCTGCTGTGATTACGAGGCGGAACTGGGCTTGTAGTTTACGCTTGGGAA----- 210  
Qy 93 AspLeuAlalysPheLysSerIleTyrValValAspLeuCysHisSerLeuGluVal 112  
Db 211 -----AGATTCAAAGAACTGATCATTCGCGATGCTTCAGAAACTATGCTGAAATG 261  
Qy 113 AlalysLysLysAlalysAlalysGlyTyrLysAsnValGlnValGluAlaAspAla 132  
Db 262 GCAGAGAAAGATCCAGCAGCAGATCTGAAAAATGTCGAGACGATCCAGCTGATGCT 321  
Qy 133 ---CysGlnPheAlaProGluGlyThrAlaThrLeuIleThrPheSerTyrSerLeu 151  
Db 322 TCTGTGCAATTT-----CCTGCTGTCAAGCTAATCTCATCTCTTATCATTGTACTT 375  
Qy 152 ThrMetIleProPheHisAsnValIleAspGlnAlaCysSerTyrLeuSerGlnAsp 171  
Db 376 TTACACATTCCTGATACGAGAGATATTCTACAAACTTTTATGAATTTTAGTCTCTGGC 435  
Qy 172 GlyLeuValGlyValAlaAspPhe 179  
Db 436 GGTCATTTGATCATCTGAGACTTT 459  
  
RESULT 6  
US-09-439-313-373  
; Sequence 373, Application US/09439313  
; Patent No. 6329505  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Mitcham, Jennifer L.  
; APPLICANT: Harlocker, Susan Louise  
; APPLICANT: Jiang Yuqi  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Kalos, Michael



Db 645 ATCCAGGAAGATGAATGTCGTTAATGTTGCTGGAACATGGCACTGATCAAAATATTC 704  
 QY 360 uGluLeuTyr-GluLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrpS 380  
 Db 705 AGATGAGTATGGAATACCACTCTGCACACTACGCTATCTATAATGAAGATAAATTAATGCG 764  
 QY 380 exLysArgLeuTyrTrp 385  
 Db 765 CAAGCACTGCTCTTAT 781

RESULT 8  
 US-09-289-198-301  
 ; Sequence 301, Application US/09289198  
 ; Patent No. 6586570  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Fridakis, Tony N.  
 ; APPLICANT: Smith, John M.  
 ; APPLICANT: Reed, Steven G.  
 ; APPLICANT: Misher, Lynda  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
 ; TREATMENT OF BREAST CANCER  
 ; FILE REFERENCE: 210121.419C5  
 ; CURRENT APPLICATION NUMBER: US/09/289,198  
 ; CURRENT FILING DATE: 1999-04-09  
 ; EARLIER APPLICATION NUMBER: US 09/062,451  
 ; EARLIER FILING DATE: 1998-04-17  
 ; EARLIER APPLICATION NUMBER: US 08/991,789  
 ; EARLIER FILING DATE: 1997-12-11  
 ; EARLIER APPLICATION NUMBER: US 08/838,762  
 ; EARLIER FILING DATE: 1997-04-09  
 ; EARLIER APPLICATION NUMBER: PCT/US97/00485  
 ; EARLIER FILING DATE: 1997-01-10  
 ; EARLIER APPLICATION NUMBER: US 08/700,014  
 ; EARLIER FILING DATE: 1996-08-20  
 ; EARLIER APPLICATION NUMBER: US 08/585,392  
 ; EARLIER FILING DATE: 1996-01-01  
 ; NUMBER OF SEQ ID NOS: 312  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 301  
 ; LENGTH: 1155  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapien  
 US-09-289-198-301

Alignment Scores:  
 Pred. No.: 0.000529 Length: 1155  
 Score: 118.50 Matches: 66  
 Percent Similarity: 37.4% Conservative: 19  
 Best Local Similarity: 29.1% Mismatches: 81  
 Query Match: 3.4% Indels: 61  
 DB: 3 Gaps: 11

US-10-620-914-45 (1-648) x US-09-289-198-301 (1-1155)

QY 187 LeuProLeuArgGlnMetProTrp-----SerArgArgPhePheTrpArgSerIlePhe 204  
 Db 204 CTTCCCTGCTGCAGGGGAGTGGCAAGCTGGCGCTCTGGAGA----- 254  
 QY 205 AspIleAspAniIleAspIleGlyProGluArgAlaTyrLeuGluGlnLysLeuGlu 224  
 Db 255 -----CCACGACACTCTGCTAT-----GAA 275  
 QY 225 ArgValTrpGluGlnAsnThrGlnGlySerIleProTyrValProTyrLeuArgAlaPro 244  
 Db 276 GACACTCAGGAACAAGATGGGCAGAGTGGTGGCCACTGCTCCCTGCTCCAGGGGAG 335  
 QY 245 TyrTyrValTrpIleGlyArgLeuProSerValGly---HisAlaLeuHisGluGluArg 263  
 Db 336 CGGCAGAGCAAGGTGGCGCTTGGGAGACTACGATCAGAGTGCCTTCATGGAGCCAG 395  
 QY 264 ValGluArgProProMetPheProThrPheLeuTyrThrGlnSerTrp-GluAsp-- 282  
 Db 396 GTACCACGTCG-----TCGAGAAGATCT 419

QY 283 -----ProGluProAspMetGluValme 290  
 Db 420 GGCAAGCTCCACAGAGCTGCCTGGTGGGTAAGTCCCAAGAAAGATCTCATCGTCAT 479  
 QY 290 t-----GluIleAsnProLysAsp-----ThrValLeuThrLeuTh 302  
 Db 480 GCTCAGGGACACTGACGTGACCAAGAAGGACAAAGCAAGAGGAGACTGCTCTACATCTGC 539  
 QY 302 rSer-----GlyGlyCysAsnAlaLeuAsnLeuValGlnGlyGlnValVa 320  
 Db 540 CTCTGCCAATGGGAATTCAGAACTAGTAAACTCTGCTGGACAGAGGATGTCACACTAA 599  
 QY 320 lSerValAspCysAsnProAlaGlnSerAlaLeuLeuGluLeuLysLysValAlaIleGl 340  
 Db 600 TGTCTTGTAC--AACAAAAGAGGACAGCTCTGATAAAG-----GCCGTACA 644  
 QY 340 nGlnLeuGluPheGluAspValTrpGlnLeuPheGlyGlyGluGlyValHisProArgIleGl 360  
 Db 645 ATGCCAGGAAGATGAATGTGCGTTAATGTTGCTGGAACTGCGCACTGATCCAAATATTC 704  
 QY 360 uGluLeuTyr-GluLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrpS 380  
 Db 705 AGATGAGTATGGAATACCACTCTGCACACTGCTATCTATAATGAAGATAAATTAATGCG 764  
 QY 380 exLysArgLeuTyrTrp 385  
 Db 765 CAAGCACTGCTCTTAT 781

RESULT 9  
 US-09-636-215-373  
 ; Sequence 373, Application US/09636215  
 ; Patent No. 6620922  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Xu, Jiangchun  
 ; APPLICANT: Dillon, Davin C.  
 ; APPLICANT: Mitcham, Jennifer L.  
 ; APPLICANT: Harlocker, Susan L.  
 ; APPLICANT: Jiang, Yuqi  
 ; APPLICANT: Henderson, Robert A.  
 ; APPLICANT: Kalos, Michael D.  
 ; APPLICANT: Fanger, Gary R.  
 ; APPLICANT: Retter, Marc W.  
 ; APPLICANT: Stolk, John A.  
 ; APPLICANT: Day, Craig H.  
 ; APPLICANT: Vedvick, Thomas S.  
 ; APPLICANT: Carter, Darrick  
 ; APPLICANT: Li, Samuel  
 ; APPLICANT: Wang, Aijun  
 ; APPLICANT: Skeiky, Yasir A.W.  
 ; APPLICANT: Hepler, William  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
 ; TREATMENT OF PROSTATE CANCER  
 ; FILE REFERENCE: 210121.42717C17  
 ; CURRENT APPLICATION NUMBER: US/09/636,215  
 ; CURRENT FILING DATE: 2000-08-10  
 ; NUMBER OF SEQ ID NOS: 852  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 373  
 ; LENGTH: 1155  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapien  
 US-09-636-215-373

Alignment Scores:  
 Pred. No.: 0.000529 Length: 1155  
 Score: 118.50 Matches: 66  
 Percent Similarity: 37.4% Conservative: 19  
 Best Local Similarity: 29.1% Mismatches: 81  
 Query Match: 3.4% Indels: 61  
 DB: 3 Gaps: 11

US-10-620-914-45 (1-648) x US-09-636-215-373 (1-1155)



GENERAL INFORMATION:  
; APPLICANT: Frudakis, Tony N.  
; APPLICANT: Smith, John M.  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Misher, Lynda  
; APPLICANT: Retter, Marc W.  
; APPLICANT: Dillon, Davin C.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
; TREATMENT AND DIAGNOSIS OF BREAST CANCER  
; FILE REFERENCE: 210121.419C6  
; CURRENT APPLICATION NUMBER: US/09/429,755A  
; CURRENT FILING DATE: 1999-10-28  
; NUMBER OF SEQ ID NOS: 315  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 301  
; LENGTH: 1155  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-09-429-755-301

Alignment Scores:  
Pred. No.: 0.000529 Length: 1155  
Score: 118.50 Matches: 66  
Percent Similarity: 37.4% Conservative: 19  
Best Local Similarity: 29.1% Mismatches: 81  
Query Match: 3.4% Indels: 61  
DB: 3 Gaps: 11

US-10-620-914-45 (1-648) x US-09-429-755-301 (1-1155)

```
QY 187 LeuProLeuArgGlnMetProTTP-----SerArgArgPhePheTTPArgSerIlePhe 204
Db 204 CTTCCTCCCTGCTGCAGGGGGAGTGGCAAGACAGAGCTGGGGCGCTTCTGAGA----- 254

QY 205 AspileAspAsnIleAspileGlyProGluArgAlaTyrLeuGluGlnLysLeuGlu 224
Db 255 -----CCACGACGACTCTGCTAT-----GAA 275

QY 225 ArgValTrpGluGlnAsnThrGlnGlySerIleProTyrValProTyrLeuArgAlaPro 244
Db 276 GACACTCAGGAACAAGATGGCAAGTGGTGTGCTCCACTGTCCCTGCTGCAGGGGGAG 335

QY 245 TyrTyrValTrpIleGlyArgLeuProSerValGly---HisAlaLeuHisGluGluArg 263
Db 336 CGGCAAGACGACAGTGGGGCTTGGGAGACTAGCAGTGCCTTCATGGAGCCGAG 395

QY 264 ValGluArgProProMetPheProProThrPheLeuTyrThrGlnSerTrp-GluAsp-- 282
Db 396 GTACCACGTCCG-----TGAGAGAAGATCT 419

QY 283 -----ProGluProAspMetGluValMe 290
Db 420 GGACACAGCTCCACAGAGCTGCCTGGTGGGGTAAAGTCCCCAGAAAGGATCTCATCGTCAT 479

QY 290 t-----GluIleAsnProLysAsp-----ThrValLeuThrLeuTh 302
Db 480 GCTCAGGACGACTGACGTGACAGAGAGGACAAAGAGGAGTCTCTACATCTGGC 539

QY 302 rSer-----GlyGlyCysAsnAlaLeuAsnLeuValGlnGlyAlaGlyGlnValVa 320
Db 540 CTCTGCCAATGGGAATTCAGAAAGTAGTAAAACTCTCTGTGGACAGACGATGCACTTAA 599

QY 320 lSerValAspCysAsnProAlaGlnSerAlaLeuLeuGluLeuLysValAlaIleGl 340
Db 600 TGCTCTTGAC---AACAAAAAGAGACGAGCTCTGATAAAG-----GCCGTACA 644

QY 340 nGlnLeuGluPheGluAspValTrpGlnLeuPheGlyGlyValHisProArgIleGl 360
Db 645 ATGCAGGACAGATGAATGTGGTTAATGTGCTGGACATGGCACTGATCCAAATATTC 704

QY 360 uGluLeuTyr-GluLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrps 380
Db 705 AGATGAGTATGGAATACCACTCTGCACTACGCTATCTATAATGAAGATAAATTAATGGC 764
```

```
QY 380 erLysArgLeuTTPTrpTyr 385
Db 765 CAAGCACTGCTCTTAT 781

RESULT 12
US-09-679-426-373
; Sequence 373, Application US/09679426
; Patent No. 679515
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqui
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TREATMENT OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C20
; CURRENT APPLICATION NUMBER: US/09/679,426
; CURRENT FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 895
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 373
; LENGTH: 1155
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-679-426-373
```

Alignment Scores:  
Pred. No.: 0.000529 Length: 1155  
Score: 118.50 Matches: 66  
Percent Similarity: 37.4% Conservative: 19  
Best Local Similarity: 29.1% Mismatches: 81  
Query Match: 3.4% Indels: 61  
DB: 3 Gaps: 11

US-10-620-914-45 (1-648) x US-09-679-426-373 (1-1155)

```
QY 187 LeuProLeuArgGlnMetProTTP-----SerArgArgPhePheTTPArgSerIlePhe 204
Db 204 CTTCCTCCCTGCTGCAGGGGGAGTGGCAAGACAGAGCTGGGGCGCTTCTGAGA----- 254

QY 205 AspileAspAsnIleAspileGlyProGluArgAlaTyrLeuGluGlnLysLeuGlu 224
Db 255 -----CCACGACGACTCTGCTAT-----GAA 275

QY 225 ArgValTrpGluGlnAsnThrGlnGlySerIleProTyrValProTyrLeuArgAlaPro 244
Db 276 GACACTCAGGAACAAGATGGCAAGTGGTGTGCTCCACTGTCCCTGCTGCAGGGGGAG 335

QY 245 TyrTyrValTrpIleGlyArgLeuProSerValGly---HisAlaLeuHisGluGluArg 263
Db 336 CGGCAAGACGACAGTGGGGCTTGGGAGACTAGCAGTGCCTTCATGGAGCCGAG 395

QY 264 ValGluArgProProMetPheProProThrPheLeuTyrThrGlnSerTrp-GluAsp-- 282
Db 396 GTACCACGTCCG-----TGAGAGAAGATCT 419

QY 283 -----ProGluProAspMetGluValMe 290
Db 420 GGACACAGCTCCACAGAGCTGCCTGGTGGGGTAAAGTCCCCAGAAAGGATCTCATCGTCAT 479
```

Qy 290 t-----GluLeuLeuProLysAsp-----ThrValLeuThrLeuTh 302  
Db 480 GCTCAGGACACTGACGTCGACAAAGAGGCAAGCAAAAGAGGACTGCTCTACATCTGGC 539  
Qy 302 rSer-----GlyGlyCysAsnAlaLeuAsnLeuValGlnGlyAlaGlyGlnValva 320  
Db 540 CTCTGCCAATGGGAATTCAGAGTAGTAAACTCTCTGTCGACAGAGCATGTCAACTTAA 599  
Qy 320 lSerValAspCysAsnProAlaGlnSerAlaLeuLeuGluLeuLysValAlaIleG1 340  
Db 600 TGCTCTTGAC---AACAAAAGAGGACAGCTCTGATAAG-----GCCGTACA 644  
Qy 340 nGlnLeuGluPheGluAspValTrpGlnLeuPheGlyGluGlyValHisProArgIleG1 360  
Db 645 ATGCAGGAAGATGAATGTGGTTAATGTGTCGGAACATGGCACTGATCCAAATATTC 704  
Qy 360 uGluLeuTyr-GluLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrpS 380  
Db 705 AGATGAGTATGGAATACCACTCTGCACATCGCTATCTATAATGAAGATAAATAATGGC 764  
Qy 380 erLysArgLeuTrpTyr 385  
Db 765 CAAAGCACTGCTCTTAT 781

## RESULT 13

US-09-759-143-373

; Sequence 373, Application US/09759143

; Patent No. 6800746

; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun

; APPLICANT: Dillon, Davin C.

; APPLICANT: Mitcham, Jennifer L.

; APPLICANT: Harlocker, Susan L.

; APPLICANT: Jiang, Yuqui

; APPLICANT: Henderson, Robert A.

; APPLICANT: Kalos, Michael D.

; APPLICANT: Fanger, Gary R.

; APPLICANT: Retter, Marc W.

; APPLICANT: Stolk, John A.

; APPLICANT: Day, Craig H.

; APPLICANT: Vedvick, Thomas S.

; APPLICANT: Carter, Darrick

; APPLICANT: Li, Samuel

; APPLICANT: Wang, Aijun

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Hepler, William

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

; FILE REFERENCE: 210121.427C23

; CURRENT APPLICATION NUMBER: US/09/759,143

; CURRENT FILING DATE: 2001-01-12

; NUMBER OF SEQ ID NOS: 934

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 373

; LENGTH: 1155

; TYPE: DNA

; ORGANISM: Homo sapien

US-09-759-143-373

## Alignment Scores:

Pred. No.:	0.000529	Length:	1155
Score:	118.50	Matches:	66
Percent Similarity:	37.4%	Conservative:	19
Best Local Similarity:	29.1%	Mismatches:	81
Query Match:	3.4%	Indels:	61
DB:	3	Gaps:	11

US-10-620-914-45 (1-648) x US-09-759-143-373 (1-1155)

Qy 187 LeuProLeuArgGlnMetProTrp-----SerArgArgPhePheTrpArgSerIlePhe 204  
Db 204 CTTCCCTGCTGCAGGGGAGTGCAGAACGTTGGCGCTTCTGGAGA----- 254

Qy 205 AspIleAspAsnIleAspIleGlyProGluArgAlaTyrLeuGluGlnLysLeuGlu 224  
Db 255 -----CCACGACGACTCTGCTAT-----GAA 275  
Qy 225 ArgValTrpGluGlnAsnThrGlnGlySerIleProTyrValProTrpLeuArgAlaPro 244  
Db 276 GACACTCAGGAAACAAGATGGGCAAGTGGTGTGCTGCTCTCCCTCGCGAGGGGAG 335  
Qy 245 TyrTyrValTrpIleGlyArgLeuProSerValGly---HisAlaLeuHisGluGluArg 263  
Db 336 CGCAAGAGCAAGGTGGCGCTTGGGGAGACTACGATGACAGTGCCTTCTATGAGGCCAG 395  
Qy 264 ValGluArgProMetPheProProThrPheLeuTyrThrGlnSerTrp-GluAsp-- 282  
Db 396 GTACCACGTCG-----TGGAGAAGATCT 419  
Qy 283 -----ProGluProAspMetGluValme 290  
Db 420 GGACAAGTCCACAGAGTGCCTGGTGGGTAAAGTCCCAAGAAAGATCTCATCGCAT 479  
Qy 290 t-----GluLeuAsnProLysAsp-----ThrValLeuThrLeuTh 302  
Db 480 GCTCAGGACACTGACGTCGACAAAGAGGCAAGCAAGAGGACTGCTCTACATCTGGC 539  
Qy 302 rSer-----GlyGlyCysAsnAlaLeuAsnLeuValGlnGlyAlaGlyGlnValva 320  
Db 540 CTCTGCCAATGGGAATTCAGAGTAGTAAACTCTCTGTCGACAGAGCATGTCAACTTAA 599  
Qy 320 lSerValAspCysAsnProAlaGlnSerAlaLeuLeuGluLeuLysValAlaIleG1 340  
Db 600 TGCTCTTGAC---AACAAAAGAGGACAGCTCTGATAAG-----GCCGTACA 644  
Qy 340 nGlnLeuGluPheGluAspValTrpGlnLeuPheGlyGluGlyValHisProArgIleG1 360  
Db 645 ATCCAGGAGATGAATGTGGTTAATGTGTCGGAACATGGCACTGATCCAAATATTC 704  
Qy 360 uGluLeuTyr-GluLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrpS 380  
Db 705 AGATGAGTATGGAATACCACTCTGCACATCGCTATCTATAATGAAGATAAATAATGGC 764  
Qy 380 erLysArgLeuTrpTyr 385  
Db 765 CAAAGCACTGCTCTTAT 781

## RESULT 14

US-09-651-236-373

; Sequence 373, Application US/09651236

; Patent No. 6818751

; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun

; APPLICANT: Dillon, Davin C.

; APPLICANT: Mitcham, Jennifer L.

; APPLICANT: Harlocker, Susan L.

; APPLICANT: Jiang, Yuqui

; APPLICANT: Henderson, Robert A.

; APPLICANT: Kalos, Michael D.

; APPLICANT: Fanger, Gary R.

; APPLICANT: Retter, Marc W.

; APPLICANT: Stolk, John A.

; APPLICANT: Vedvick, Thomas S.

; APPLICANT: Carter, Darrick

; APPLICANT: Li, Samuel

; APPLICANT: Wang, Aijun

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Hepler, William

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

; FILE REFERENCE: 210121.42718C18

; CURRENT APPLICATION NUMBER: US/09/651,236

; CURRENT FILING DATE: 2000-08-29

; NUMBER OF SEQ ID NOS: 865

; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 373  
; LENGTH: 1155  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-09-651-236-373

Alignment Scores:  
Pred. No.: 0.000529 Length: 1155  
Score: 118.50 Matches: 66  
Percent Similarity: 37.4% Conservative: 19  
Best Local Similarity: 29.1% Mismatches: 81  
Query Match: 3.4% Indels: 61  
DB: 3 Gaps: 11

US-10-620-914-45 (1-648) x US-09-651-236-373 (1-1155)

```
Qy 187 LeuProLeuArgGlnMetProTrp-----SerArgArgPhePheTrpArgSerIlePhe 204
Db 204 CTTCCCTGCTGCAGGGGGAGTGGCAAGAGCAACGTGGGGCTTCTGGAGA-----254
Qy 205 AspileaspnleaspilleGlyProGluArgAlaTyrlLeuGluGlnLysLeuGlu 224
Db 255 -----CCACGACGACTCTGCTAT-----GAA 275
Qy 225 ArgValTrpGluGlnAenThrGlnGlySerIleProTyrlValProTrpLeuArgAlaPro 244
Db 276 GACACTCAGGACACAGATGGCAAGTGGTGTGCTGCCACTGCTTCCCTGCTGCAGGGGGAG 335
Qy 245 TyrlValTrpIleGlyArgLeuProSerValGly---HisAlaLeuHisGluGluArg 263
Db 336 CGGCAAGACGCAAGTGGGGCTTGGGAGAGACTACGATGACAGTGCCTTTCATGAGCCGAG 395
Qy 264 ValGluArgProProMetPheProProThrPheLeuTyrlThrGlnSerTrp-GluAsp-- 282
Db 396 GTACCACGTCGG-----TGAGAAGATCT 419
Qy 283 -----ProGluProAspMetGluValMe 290
Db 420 GGACACGCTCCACAGAGCTGCTGGTGGGTAAAGTCCCGAAGAGATCTCATCGTCAT 479
Qy 290 t-----GluLeuAenProLysAsp-----ThrValLeuThrLeuTh 302
Db 480 GCTCAGGACACTGACCTGACCAAGAGAGGACAAAGAGGAGGACTGCTTACATCTGGC 539
Qy 302 rSer-----GlyGlyCysAsnAlaLeuAenLeuValGlnGlyAlaGlyGlnValVa 320
Db 540 CTCTGCCAATGGGAATTCAGAAAGTAGTAAACCTCTCTGGACAGACGATGTCAACTTAA 599
Qy 320 lSerValAspCysAsnProAlaGlnSerAlaLeuLeuGluLeuLysLysValAlaIleGl 340
Db 600 TGTCCTTGAC---AACAAAAAGAGGACAGCTCTGATAAAG-----GCCGTACA 644
Qy 340 nGlnLeuGluPheGluAspValTrpGlnLeuPheGlyGluGlyValHisProArgIleGl 360
Db 645 ATGCCAGGACATGAATGCTGTTAATGTTGCTGGACATGGCACTGATCAATAATTC 704
Qy 360 uGluLeuTyrl-GluLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrpS 380
Db 705 AGATGAGTATGGAAATACCACTCTGCACTACGCTATCTATAATGAAGATAAATTAATGGC 764
Qy 380 erLysArgLeuTrpTyrl 385
Db 765 CAAGCAGCTGCTCTTAT 781
```

## RESULT 15

US-09-699-295-301  
; Sequence 301, Application US/09699295  
; Patent No. 5828431  
; GENERAL INFORMATION:  
; APPLICANT: Frudakis, Tony N.  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Smith, John M.

; APPLICANT: Misher, Linda E.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Retter, Marc W.  
; APPLICANT: Wang, Aijun  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Harlocker, Susan L.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
; FILE REFERENCE: 210121.419C10  
; CURRENT APPLICATION NUMBER: US/09/699,295  
; CURRENT FILING DATE: 2000-10-26  
; NUMBER OF SEQ ID NOS: 326  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 301  
; LENGTH: 1155  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-09-699-295-301

Alignment Scores:  
Pred. No.: 0.000529 Length: 1155  
Score: 118.50 Matches: 66  
Percent Similarity: 37.4% Conservative: 19  
Best Local Similarity: 29.1% Mismatches: 81  
Query Match: 3.4% Indels: 61  
DB: 3 Gaps: 11

US-10-620-914-45 (1-648) x US-09-699-295-301 (1-1155)

```
Qy 187 LeuProLeuArgGlnMetProTrp-----SerArgArgPhePheTrpArgSerIlePhe 204
Db 204 CTTCCCTGCTGCAGGGGGAGTGGCAAGAGCAACGTGGGGCTTCTGGAGA-----254
Qy 205 AspileaspnleaspilleGlyProGluArgAlaTyrlLeuGluGlnLysLeuGlu 224
Db 255 -----CCACGACGACTCTGCTAT-----GAA 275
Qy 225 ArgValTrpGluGlnAenThrGlnGlySerIleProTyrlValProTrpLeuArgAlaPro 244
Db 276 GACACTCAGGACACAGATGGCAAGTGGTGTGCTGCCACTGCTTCCCTGCTGCAGGGGGAG 335
Qy 245 TyrlValTrpIleGlyArgLeuProSerValGly---HisAlaLeuHisGluGluArg 263
Db 336 CGGCAAGACGCAAGTGGGGCTTGGGAGAGACTACGATGACAGTGCCTTTCATGAGCCGAG 395
Qy 264 ValGluArgProProMetPheProProThrPheLeuTyrlThrGlnSerTrp-GluAsp-- 282
Db 396 GTACCACGTCGG-----TGAGAAGATCT 419
Qy 283 -----ProGluProAspMetGluValMe 290
Db 420 GGACACGCTCCACAGAGCTGCTGGTGGGTAAAGTCCCGAAGAGATCTCATCGTCAT 479
Qy 290 t-----GluLeuAenProLysAsp-----ThrValLeuThrLeuTh 302
Db 480 GCTCAGGACACTGACCTGACCAAGAGAGGACAAAGAGGAGGACTGCTTACATCTGGC 539
Qy 302 rSer-----GlyGlyCysAsnAlaLeuAenLeuValGlnGlyAlaGlyGlnValVa 320
Db 540 CTCTGCCAATGGGAATTCAGAAAGTAGTAAACCTCTCTGGACAGACGATGTCAACTTAA 599
Qy 320 lSerValAspCysAsnProAlaGlnSerAlaLeuLeuGluLeuLysLysValAlaIleGl 340
Db 600 TGTCCTTGAC---AACAAAAAGAGGACAGCTCTGATAAAG-----GCCGTACA 644
Qy 340 nGlnLeuGluPheGluAspValTrpGlnLeuPheGlyGluGlyValHisProArgIleGl 360
Db 645 ATGCCAGGACATGAATGCTGTTAATGTTGCTGGACATGGCACTGATCAATAATTC 704
Qy 360 uGluLeuTyrl-GluLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrpS 380
Db 705 AGATGAGTATGGAAATACCACTCTGCACTACGCTATCTATAATGAAGATAAATTAATGGC 764
```



Qy 380 erLysArgLeuTrpTyr 385  
Db 765 CAAGCACTGCTCTTAT 781

Search completed: March 14, 2006, 07:38:00  
Job time : 324 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7

Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: March 14, 2006, 05:00:51 ; Search time 1369 Seconds

(without alignments)

3914.213 Million cell updates/sec

Title: US-10-620-914-45

Perfect score: 3463

Sequence: 1 MGSGRDRPASYTKNFSLE.....RVNYSSTFVARRKGAKKDN 648

Scoring table:

BLOSUM62	Xgapop 10.0 , Xgapext 0.5
	Ygapop 10.0 , Ygapext 0.5
	Fgapop 6.0 , Fgapext 7.0
	Delop 6.0 , Delext 7.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

```

-MODEL=frame+ p2n.model -DEV=xlh
-O=/abs/ABSSWEB.spool/US10620914/runat.13032006.102049.6893/app_query.fasta_1
-DB=Published Applications_NA_Main -QFMT=fastap -SUFFIX=p2n.rnpbm
-MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0 -LIST=45 -MODE=LOCAL -OUTFMT=ptc -NORM=ext
-MATRIX=biosum62 -TRANS=human40.cdi -ALIGN=15 -DOCL=LOCAL -DEV_TIMEOUT=120
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -DOCL=LOCAL -OUTFMT=ptc -NORM=ext
-HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abss05h
-USER=US10620914 @CGN 1.1.1026 @runat.13032006.102049.6893 -NCPU=6 -ICPU=3
-NO MMAP -NEG_SCORES=0 -WAIT -BSPLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=0.5 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

```

Database : Published Applications\_NA\_Main:

1:	/cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq*
2:	/cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq*
3:	/cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq*
4:	/cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq*
5:	/cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq*
6:	/cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq*
7:	/cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq*
8:	/cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq*
9:	/cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq*
10:	/cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3463	100.0	1947	7	US-10-620-914-44
2	2838	82.0	5242	7	US-10-620-914-43
3	891	25.7	2259	8	US-10-741-849-6021
4	850.5	24.6	2727	7	US-10-620-914-49
5	804	23.2	3427	7	US-10-620-914-48
6	379.5	11.0	1252	5	US-10-118-495-1
7	379.5	11.0	1252	7	US-10-620-914-1

8	329	9.5	1248	5	US-10-118-495-22	Sequence 22, Appl
9	329	9.5	1248	7	US-10-620-914-22	Sequence 22, Appl
10	326.5	9.4	1251	5	US-10-118-495-28	Sequence 28, Appl
11	326.5	9.4	1251	7	US-10-620-914-28	Sequence 28, Appl
12	292.5	8.4	1251	5	US-10-118-495-32	Sequence 32, Appl
13	292.5	8.4	1251	7	US-10-620-914-32	Sequence 32, Appl
14	128	3.7	4246	6	US-10-085-959-12	Sequence 12, Appl
15	123	3.6	7119	8	US-10-684-141-57	Sequence 57, Appl
16	123	3.6	7119	9	US-10-810-486-57	Sequence 57, Appl
17	120	3.5	666	5	US-10-118-495-34	Sequence 34, Appl
18	120	3.5	666	7	US-10-620-914-34	Sequence 34, Appl
19	119	3.4	6912	8	US-10-335-053-309	Sequence 309, Appl
20	119	3.4	6912	8	US-10-684-141-53	Sequence 53, Appl
21	119	3.4	6912	9	US-10-810-486-53	Sequence 53, Appl
22	119	3.4	6912	9	US-10-887-553A-400	Sequence 400, Appl
23	118.5	3.4	1155	3	US-09-825-301-5	Sequence 5, Appl
24	118.5	3.4	1155	3	US-09-759-143-373	Sequence 373, Appl
25	118.5	3.4	1155	3	US-09-780-669-373	Sequence 373, Appl
26	118.5	3.4	1155	3	US-09-810-936-301	Sequence 301, Appl
27	118.5	3.4	1155	3	US-09-822-827-373	Sequence 373, Appl
28	118.5	3.4	1155	3	US-09-429-755-301	Sequence 301, Appl
29	118.5	3.4	1155	3	US-09-924-400-301	Sequence 301, Appl
30	118.5	3.4	1155	3	US-09-895-793-373	Sequence 373, Appl
31	118.5	3.4	1155	3	US-09-895-814-373	Sequence 373, Appl
32	118.5	3.4	1155	5	US-10-012-896-373	Sequence 373, Appl
33	118.5	3.4	1155	5	US-10-010-940-373	Sequence 373, Appl
34	118.5	3.4	1155	6	US-10-212-679-301	Sequence 301, Appl
35	118.5	3.4	1155	6	US-10-144-678A-373	Sequence 373, Appl
36	118.5	3.4	1155	6	US-10-033-527-5	Sequence 5, Appl
37	118.5	3.4	1155	6	US-10-294-025-373	Sequence 373, Appl
38	118.5	3.4	1155	7	US-10-079-137B-301	Sequence 301, Appl
39	118.5	3.4	1185	3	US-09-924-400-335	Sequence 335, Appl
40	118.5	3.4	1185	6	US-10-212-679-335	Sequence 335, Appl
41	118.5	3.4	1185	7	US-10-079-137B-335	Sequence 335, Appl
42	118.5	3.4	1383	3	US-09-905-673-64	Sequence 64, Appl
43	118.5	3.4	1383	6	US-10-096-319-64	Sequence 64, Appl
44	118.5	3.4	1434	3	US-09-905-673-65	Sequence 65, Appl
45	118.5	3.4	1434	6	US-10-096-319-65	Sequence 65, Appl

## ALIGNMENTS

## RESULT 1

```

US-10-620-914-44
; Sequence 44, Application US/10620914
; Publication No. US20040093639A1
; GENERAL INFORMATION:
; APPLICANT: Benning, Christoph
; APPLICANT: Riekhof, Wayne
; TITLE OF INVENTION: Klug, Rouven
; FILE REFERENCE: MSU-07769
; CURRENT APPLICATION NUMBER: US/10/620,914
; CURRENT FILING DATE: 2003-07-16
; PRIOR APPLICATION NUMBER: 10/118,495
; PRIOR FILING DATE: 2002-04-08
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 44
; TYPE: DNA
; LENGTH: 1947
; ORGANISM: Chlamydomonas reinhardtii
US-10-620-914-44

```

Alignment Scores:

Pred. No.:	0	Length:	1947
Score:	3463.00	Matches:	648
Percent Similarity:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	100.0%	Indels:	0
DB:	7	Gaps:	0

US-10-620-914-45 (1-648) x US-10-620-914-44 (1-1947)

QY 1 MetGlySerGlyArgAspGlyArgProAlaSerTyrThrLysLysAsnPheSerLeuGlu 20  
DB 1 ATGGGGTGGGTCTGACGGCGGCTTGGAGCTACACCAAGAGAACTTCTCCCTGGAG 60  
QY 21 LysLeuLysLeuSerSerMetLysAspAspLeuThrValLeuArgHisMetTyrPheGly 40  
DB 61 AAGCTCAAGCTCAGCAGCATGAAGGATGACCTGACCGTTCTGGCCCATATGTGTTCGGC 120  
QY 41 SerLysLysGlyAspAspHisAlaAlaArgLeuGluSerPheTyrGlyProGlnAlaAla 60  
DB 121 AGCAAGAGCGGATGATCATCGCTGCTCGCTGGAGAGCTTCTACGGGCGCCCAAGCGCGCT 180  
QY 61 AlaPheAlaAlaArgLeuAlaGluArgSerAsnLeuLeuTyrValAspLeuGlyGly 80  
DB 181 GCCTTTGCTGCCCGCTGGCGGAGCGCTCGAACCTCATCTGGGTGTGACCTGGGTGGTGGC 240  
QY 81 ThrGlyGluAsnValAspMetMetAlaAspTyrIleAspLeuAlaLysPheLysSerIle 100  
DB 241 ACTGGGAGAAATGTCGATATGATGCTGATTTACATCGACCTGGCGAAGTTTCAAGTCCATC 300  
QY 101 TyrValValAspLeuCysHisSerLeuCysGluValAlaLysLysLysAlaLysAlaLys 120  
DB 301 TACGTGCTCGACCTGTGCCACTCGCTGTGGAGGTGGCCCAAGAGAGAGGCGCAAG 360  
QY 121 GlyTTrpLysAsnValGlnValValGluAlaAspAlaCysGlnPheAlaProProGluGly 140  
DB 361 GCCTGGGAAGAAATGTCAGGTGCTGGAGGCGACCTTGGCAATTTGGCGCCCTTGAGGGC 420  
QY 141 ThrAlaThrLeuIleThrPheSerTyrSerLeuThrMetIleProProPheHisAsnVal 160  
DB 421 ACCGCGACGCTCATCCTCTCTACTCTGCTACGATGATTTCCACCGTTTCCCAACGTC 480  
QY 161 IleAspGlnAlaCysSerTyrLeuSerGlnAspGlyLeuValGlyValAlaAlaAspPheTyr 180  
DB 481 ATCGACGAGGCTTGCTGCTACCTGCTCCCAAGACGGCTGTGGTGGCGCTTTCCTAC 540  
QY 181 ValSerGlyLysTyrAspLeuProLeuArgGlnMetProTyrSerArgAspPhePheTyr 200  
DB 541 GTGAGCGGCAAGTACGACCTGCTCCCTGCGGAGATGCTGCTGGTGGCGGCTTCTTCGG 600  
QY 201 ArgSerIlePheAspIleAspAsnIleAspIleGlyProGluArgArgAlaTyrLeuGlu 220  
DB 601 CGATCGATCTTCGACATCGACAACATTTGACATCGGCGCGGCGCGCTACCTGGAG 660  
QY 221 GlnLysLeuGluArgValTyrGluGlnAsnThrGlnLysSerIleProTyrValProTyr 240  
DB 661 CAGAAGCTGGAGCGCGTGTGGGAGCAGAACACCCAGGGTTTCGATCCCTACGTGCGGTGG 720  
QY 241 LeuArgAlaProTyrTyrValTyrIleGlyArgLeuProSerValGlyHisAlaLeuHis 260  
DB 721 CTGGCGCCCTTACTACGTGTGATTTGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTTGCAC 780  
QY 261 GluGluArgValGluArgProProMetPheProProThrPheLeuTyrThrGlnSerTyr 280  
DB 781 GAGGAGCGGTGGAGCGCGCCCATGTTCCCGCCCATCTTCTGTACACGACGTCTGG 840  
QY 281 GluAspProGluProAspMetGluValMetGluIleAsnProLysAspThrValLeuThr 300  
DB 841 GAGGACCCCGAGCGGATATGAGGTGATGAGATCAACCCCAAGAGACACGCGTCTGACC 900  
QY 301 LeuThrSerGlyGlyCysAsnAlaLeuAsnLeuValGlnGlyAlaGlyGlnValVal 320  
DB 901 CTGACTAGCGCGCTGCATGCTCCCTGACCTGTGTGGTGGAGGGCGCGGCGGCTGGT 960  
QY 321 SerValAspCysAsnProAlaGlnSerAlaLeuLeuGluLeuLysLysValAlaIleGln 340  
DB 961 TCGGTGGACTGCAACCCCGCGAGTCGCGCTTCTGGAGCTGAAGAGAGGTGGCCATTTCAG 1020  
QY 341 GlnLeuGluPheGluAspValTyrGlnLeuPheGlyGlnGlyValHisProArgIleGlu 360  
DB 1021 CAGCTGAGTTTGGAGCGTGTGGCAGCTGTTTCGGGAGGGCGTGGCAGCGCGCATTTGAG 1080

QY 361 GluLeuTyrGluLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTyrSer 380  
DB 1081 GAGCTGTACGAGAAGAGCTGGCGGCTTCTCTGTCGAACACGAGCACAACTTCTGTGTCC 1140  
QY 381 LysArgLeuTyrTyrPheGlnHisGlyLeuTyrTyrGlnGlyMetGlyLysLeuCys 400  
DB 1141 AAGCGCTCTGTGTACTTCCAGCAGCGCTGTACTACGAGCGGCGCATGGGCAAGTGTGC 1200  
QY 401 TrpValLeuGlnCysLeuAlaValValLeuGlyLeuGlyLysThrValLysArgLeuAla 420  
DB 1201 TGGGTGCTGAGTGGCTTGGCGGCTGGTGGAGCTGGGCAAGACCGTCAAGCGGCTCGCC 1260  
QY 421 AsnAlaProThrMetGluGluGlnArgArgLeuTyrAspSerAsnMetLeuIleHisPhe 440  
DB 1261 AACGCGCCACATGGAGGAGCAGCGCGTCTGTGGGACAGCAACATGCTCATCCACTTC 1320  
QY 441 ValLysAsnGlyProLysProLeuValTyrLeuPheValLysPheValSerLeuValLeu 460  
DB 1321 GTGAAGAACCGGCGCAAGCGCTGTGTGGTGTCTTCTCAAGTTCTGTAGCGCTGTGTCTC 1380  
QY 461 PheAsnLysAlaValLeuTyrPheGlyGlyValProGlyLysGlnTyrAlaLeuIle 480  
DB 1381 TTCAACAAGCCGCTGTGTGGCGGCGGCTGGCGGCAAGCAAGTACGCGCTGATC 1440  
QY 481 LysAlaAspGlyIleProIleGluAsnTyrIleAlaArgThrMetAspGlyValAlaGlu 500  
DB 1441 AAGCGGAGCGCATCCCATTTGAGAACTACATCGCGCACCATGAGCGCGTGGCGGAG 1500  
QY 501 AsnSerHisValArgLysGlnAsnTyrPheTyrTyrAsnCysLeuThrGlyLysPheLeu 520  
DB 1501 AACTCGCAGCTGGCGCAAGCAGAACTACTTCTACTACAACTGCTCACCAGTTCCTGT 1560  
QY 521 ArgAspAsnCysProThrTyrLeuArgGluAlaAlaPheAlaThrLeuLysSerGlyVal 540  
DB 1561 CGGCAACACTGGCCCACTTACCTGCGGAGCGGCTTCCCACTTCAAGAGTGGCGTG 1620  
QY 541 ValAspAsnLeuThrValSerThrAsnPhePheMetGluLeuLysAlaArgThrTyr 560  
DB 1621 GTGGCAACCTGACCGCTCTCCACCACTTCTTATGGAGGAGCTCAAAGCGCGCACCTAC 1680  
QY 561 ThrLysValIleLeuMetAspHisValAspTyrLeuAspMetProValAlaAsnGluLeu 580  
DB 1681 ACCAAGGTGATTTCTGTATGGACACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740  
QY 581 AlaGluCysLeuAlaLysGlnValAlaProGlyGlyValIleTyrArgSerAlaSer 600  
DB 1741 GCGAGTGGCTGGCCAGCAGGTGTGCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1800  
QY 601 LeuSerProTyrAlaGluLeuIleGlnLysAlaGlyPheAspValArgCysIleArg 620  
DB 1801 CTGAGCGCGCTTACGCGAGCTGATCCAGAAGCGGCTTTCGAGCTGCTGCTGCTGCTGCTGCTG 1860  
QY 621 ArgAlaThrGlnGlyTyrMetAspArgValAsnMetTyrSerSerPheTyrMetAlaArg 640  
DB 1861 CGCGCCACTCAGGCTACATGGACCGGCTCAACATGTACAGCTCTCTTCTCATGCGCGCTGCTGCTG 1920  
QY 641 ArgLysGlyAlaLysLysAspAsn 648  
DB 1921 CGGAAGGCGCGCCCAAGAGGACAAC 1944

## RESULT 2

US-10-620-914-43  
; Sequence 43, Application US/10620914  
; Publication No. US20040093639A1  
; GENERAL INFORMATION:  
; APPLICANT: Benning, Christoph  
; APPLICANT: Riekhof, Wayne  
; APPLICANT: Klug, Rouven  
; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids  
; FILE REFERENCE: WSU-0769  
; CURRENT APPLICATION NUMBER: US/10/620,914  
; CURRENT FILING DATE: 2003-07-16  
; PRIOR APPLICATION NUMBER: 10/118,495

; PRIOR FILING DATE: 2002-04-08  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 43  
; LENGTH: 5242  
; TYPE: DNA  
; ORGANISM: Chlamydomonas reinhardtii  
US-10-620-914-43

## Alignment Scores:

Pred. No.: 0 Length: 5242  
Score: 2838.00 Matches: 645  
Percent Similarity: 45.6% Conservatives: 0  
Best Local Similarity: 45.6% Mismatches: 3  
Query Match: 82.0% Indels: 771  
DB: 7 Gaps: 11

US-10-620-914-45 (1-648) x US-10-620-914-43 (1-5242)

Qy 1 MetGlySerGlyArgAspGlyAProAlaSerTyrThrLysLysAsnPheSerLeuGlu 20  
Db |||||  
Qy 501 ATGGGGTGGGTCGTGACGGCCGGCTGCGAGCTACACCAAGAAGAACTTCTCCCTGGAG 560  
Db |||||  
Qy 21 LysLeuLysLeuSerSerMetLysAspAspLeuThrValLeuArgHisMetTrpPheGly 40  
Db |||||  
Qy 561 AAGCTCAAGCTCAGCAGCATGAAGGATGACCTGACCGCTTCTGGCCATATGTGGTTCGGC 620  
Db |||||  
Qy 41 SerLysLysGlyAspAspHisAlaAlaArgLeuGluSerPheTyrGlyProGlnAlaAla 60  
Db |||||  
Qy 621 AGCAAGAGGCGCATGATACGCTGCTGCGCTGGAGAGCTTCTACGGGCCCAAGCGCT 680  
Db |||||  
Qy 61 AlaPhe----- 62  
Db |||||  
Qy 681 GCCTG-TAAGTCGGATTAACCTTTATGCTCATTTAAGTGTTCGAGTGTCTTACGGCGTG 739  
Db |||||  
Qy 62 ----- 62  
Db |||||  
Qy 740 TCTGCCCGCCAGATGATGTTTCCGGTCCGGTTCCTCTGGGGTCCGAGGCCCATGCTC 799  
Db |||||  
Qy 63 -----AlaAlaArgLeuAlaGluArgSerAsnLeuIleTrpValAspLeuGly 79  
Db |||||  
Qy 800 GCTGCAGTGTCTGCCCGCTGGCGAGCGCTCGAACCTCATCTGGGTGACCTGGGTGGT 859  
Db |||||  
Qy 80 GlyThrGly----- 82  
Db |||||  
Qy 860 GGCACTGGGTGAGTCGCGTGTGATCGGGAGGTGCGGATAGCCCTGCGCGCTTTGGCT 919  
Db |||||  
Qy 83 -----GluAsnVal 85  
Db |||||  
Qy 920 GCGCCTCGCGCGCATTTATGCTGATCTGCGCTTTGTCTTTGTGCGCGAGGAGATGTC 979  
Db |||||  
Qy 86 AspMetMetAlaAspTyrIleAspLeuAlaLysPheLysSerIleTyrValValAspLeu 105  
Db |||||  
Qy 980 GATATGATGCTGATTACATCGACCTGGCGAAGTTCAAGTCCATCTACGTGGTGCACCTG 1039  
Db |||||  
Qy 106 CysHisSerLeuCysGluValAlaLysLysAlaLysAlaLysGlyTrpLysAsnVal 125  
Db |||||  
Qy 1040 TGCCACTCGCTGTGCGAGGTGGCCAGAGCAAGCGCAAGGCCAAGGGCTGGAAGAATGTC 1099  
Db |||||  
Qy 126 GlnValValGluAlaAspAlaCysGlnPheAlaProProGluGlyThrAlaThrLeuIle 145  
Db |||||  
Qy 1100 CAGTGTGGAGGCGGACGCTGTCCCAATTTGCGCCCTTGGGGCCACCGGACGCTCATC 1159  
Db |||||  
Qy 146 ThrPheSerTyrSerLeuThrMet----- 153  
Db |||||  
Qy 1160 ACCTTCTCTACTCTGCTCACGAG-TGAGTTGCAACCGCGCTGCACTTGCCTACGAGATC 1218  
Db |||||  
Qy 154 -----IlePro 155  
Db |||||  
Qy 1219 CATCCACCCACACAGTTTCATCCCTCTCACCCCGCGCTTTTGTGTTGCGAGTATCCA 1278  
Db |||||  
Qy 156 PropHHisAsnValIleAspGlnAlaCysSerTyrLeuSerGlnAspGlyLeuValGly 175  
Db |||||

Db 1279 CCGTTCACAAAGTCATCGACCAAGGCTTGTCTGTACTCTGCCAAGAGCGCCTGTGGGC 1338  
Qy ValAlaAspPheTyrValSerGlyLysTyrAspLeuProLeuArgGlnMetProTrpSer 195  
Db |||||  
Qy 1339 GTTGGCGACTTCTACGTGAGCGCAAGTAGACCTGCCCTCGCCAGATGCCCTGTGTCG 1398  
Qy ArgArgPhePheTrp----- 200  
Db |||||  
Qy 1399 CGCCGTTTCTTCTGGCGGTGAGTTACCAGGCGCGGCTACATCTCTCAAGCAAGCGTA 1458  
Qy ----- 200  
Db ----- 200  
Qy 1459 GCTGACGGGGCACAGAACGCTTAGGGCCCGGGTGGCTTCTGGGTAGTGGCAGGA 1518  
Qy ----- 200  
Db ----- 200  
Qy 1519 TATCGCAAGGCCCTTCAGTACCCGAGGTACCATGTTGATCTGTCAAGCGAGCTCATCAC 1578  
Qy ----- 200  
Db ----- 200  
Qy 1579 GGTGCTGACTGAGGACAGTCTCGAGTCTCGGTCGTCTAGGGCTTTCAGCACATCACA 1638  
Qy ----- 200  
Db ----- 200  
Qy 1639 GCGAGTCTCCGNAATCGCGTGTCTGAGCGGTGTATGCTCGGGACAGCACCATGCAACGT 1698  
Qy ----- 200  
Db ----- 200  
Qy 1699 GCGGCGTCTTGACTTTCATCGCTCTCTCGGGCTCCCTCGGTTTCGCTCAGACACGT 1758  
Qy -----ArgSerIlePheAspAsn 208  
Db |||||  
Qy 1759 CTCCGCTGCCCTGTGTCTGCTGCTGCTGTGCGATCGATCTTCGACATCGCAAC 1818  
Qy ----- 208  
Db ----- 208  
Qy 209 IleAspIleGlyProGluArgAlaTyrLeuGluGlnLysLeuGluArgValTrpGlu 228  
Db |||||  
Qy 1819 ATTGATCGGCCCGCGCGCTACCTGAGCAGAGCTGAGCGCGTGTGGAG 1878  
Qy ----- 231  
Db ----- 231  
Qy 229 GlnAsnThr----- 231  
Db |||||  
Qy 1879 CAGAACACCCAGGTGTGTCGGTGTGACACCTGGCTGTGCAAGCCCAAGTAAGTTTG 1938  
Qy ----- 231  
Db ----- 231  
Qy 1939 CTCATCAGCCCTCGACGACCCGCGGTGCGCGGATATTCTGCGCAGAGCGCTCCGCA 1998  
Qy -----GlnGlySerIlePro 236  
Db |||||  
Qy 1999 CTGTTGGTTAATCCACAGTCGCGCTGCTTCTCTGCGTGTTCAGGGTTTCGATCCCC 2058  
Qy ----- 256  
Db ----- 2118  
Qy 237 TyrValProTrpLeuArgAlaProTyrTyrValTrpIleGlyArgLeuProSerValGly 256  
Db |||||  
Qy 2059 TACGTGCGGTGCTGCGCGCCCTACTACGTGTGATTTGGCGCTGCCAGGTTGGC 2118  
Qy ----- 257  
Db ----- 2177  
Qy 2119 CG-TGAGTCGCTGCGCGGACCTCGTCCGATCAGGACACATGTTACGGGGCGCTCA 2177  
Qy ----- 257  
Db ----- 2337  
Qy 2178 GTGAAGCGTATCGTATCGTTGCGATCGTACTGTCAAGCGGTACCAACAGCGCGCTCCCTTG 2337  
Qy ----- 257  
Db ----- 2238  
Qy 2238 CGGTTGGAATTTGTTGGCGCAAGCATGAGCGGTGCGCGATGCCCACTCGCGTTGAC 2297  
Qy -----AlaLeuHisGluArgValGluArgProProMetPheProThrPhe 274  
Db |||||  
Qy 2298 CTTGACAGACCGCTGACGAGGAGCGGTGAGCGCGCCCATGTTCCGCCACCTTC 2357  
Qy ----- 288  
Db ----- 2358  
Qy 275 LeuTyrThrGlnSerTrpGluAspProGluProAspMetGlu----- 288  
Db |||||  
Qy 2358 CTGTACACGAGTCTGTGGGAGGACCCCGAGCGGATATGAGAGGTGAGCGGCTCAGCGCTGA 2417  
Db ----- 2417



Db 4578 TGTTCGATACGAGGACGCTAGCGTCAGGGCGCAACAGCGCTGGCGCACCGGAAA 4637  
 QY 634 -----Ser 634  
 Db 4638 TCTGCGTTGCTTCGATGTATACCTCGATCCAACTCTTGCTGGTTCTCCMATGGCAGC 4697  
 QY 635 SerPheTyrMetAlaArgArgGlyGlyAlaLysAlaLysAspAsn 648  
 Db 4698 TCCTTCTACATGCGCGCGGAGGCGCCCAAGAGGACAAAC 4739

## RESULT 3

US-10-741-849-6021  
 ; Sequence 6021, Application US/10741849  
 ; Publication No. US20050019931A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Roemer, Terry  
 ; APPLICANT: Jiang, Bo  
 ; APPLICANT: Boone, Charles  
 ; APPLICANT: Bussey, Howard  
 ; TITLE OF INVENTION: Nucleic Acids Encoding Anti-fungal Drug Targets and Methods of  
 ; FILE REFERENCE: 10182-023-999  
 ; CURRENT APPLICATION NUMBER: US/10741,849  
 ; CURRENT FILING DATE: 2003-12-19  
 ; PRIOR APPLICATION NUMBER: US 60/434,832  
 ; PRIOR FILING DATE: 2002-12-19  
 ; NUMBER OF SEQ ID NOS: 8000  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 6021  
 ; LENGTH: 2259  
 ; TYPE: DNA  
 ; ORGANISM: Candida albicans  
 US-10-741-849-6021

## Alignment Scores:

Prod. No.: 3,81e-99 Length: 2259  
 Score: 891.00 Matches: 229  
 Percent Similarity: 46.6% Conservative: 98  
 Best Local Similarity: 32.7% Mismatches: 232  
 Query Match: 25.7% Indels: 142  
 DB: 8 Gaps: 21

US-10-620-914-45 (1-648) x US-10-741-849-6021 (1-2259)

QY 51 LeuGluSerPheTyrGlyProGlnAlaAlaAlaPhe----- 62  
 Db 208 TTAGAGTCATCTACAGAATCAAGCTCATATTTATGATACACTAGAGAAATTTTGTG 267  
 QY 63 -----AlaAlaArgLeuAlaGluArgSerAsnLeu 72  
 Db 268 AAAGGTAGACAAGATGCTTAGATTAGCTATTTCCCAATTTCCCAAAAGAAAGATCTT 327  
 QY 73 IleTyrValAspLeuGlyGlyThrGlyGluAsnValAspMetMetAlaAspTyrIle 92  
 Db 328 ATTTGGATTCATATTTGGTGGAACTGGTTCCAAATATTGAATTCATGGATGAATTAGT 387  
 QY 93 AspLeuAlaLys---PheLysSerIleTyrValValAspLeuCysHisSerLeuCysGlu 111  
 Db 388 AAAATATCTGAAAACCTTTAAAGCTGTTATTTGGTTGATCTTTCCCATCTTTTGTGGAA 447  
 QY 112 ValAlaLysLysLysAlaLysAlaLysGlyTyrLysAsnValGlnValValGluAlaAsp 131  
 Db 448 GTTGCTAAGCAGATTTGAGGCCCATGATGATGACAAATGTTTCATGATTTAGTTGCTGAT 507  
 QY 132 AlaCysGlnPheAlaProProGluGlyThrAlaThrLeuIleThrPheSerTyrSerLeu 151  
 Db 508 GCCTGTGATTTACTATTGATTGATGATGCTGATTTGATTGATTTTCTTATTCTATTG 567  
 QY 152 ThrMetIleProProPheHisAsnValIleAspGlnAlaCysSerTyrLeuSerGlnAsp 171  
 Db 568 TCGATGATCCAACTTTCAATGCTGCTATCGTAATGCTGTTCTTAAATTAGATATGGAA 627

QY 172 GlyLeuValGlyValAlaAspPheTyrValSer-----GlyLysTyr 185  
 Db 628 GGTATTATTATTCACCTGCTGGATTTGGTATTCAAAGCAGTGACACCTCAATGGTGGTATC 687  
 QY 186 Asp-----LeuProLeuArgGlnMetProTyrSerArgArgPhePheTyrArg 201  
 Db 688 ATATCTGTTGGTGGTGGTTTACAGGACATTCCTTGGATATTACGTAAATTTTGGAGA 747  
 QY 202 SerIlePheAspIleAspAsnIleAspIleGlyProGluArgArgAlaTyrLeuGluGln 221  
 Db 748 ATTTGGTTTGAAGCTGATAAAGTGTTTTGGATCTTCAAGAAGAAACATTTTGGAAATAT 807  
 QY 222 LysLeuGluArgValTyrGluGlnAsnThrGln-----GlySerIleProTyr 237  
 Db 808 AAATTTGGTACCGTCAAAATCTTTGAATTCATACCAAGGCTTTGGGTAAATC----- 861  
 QY 238 ValProTyrLeuArgAlaProTyrTyrValTyrIleGlyArgLeuProSerValGlyHis 257  
 Db 862 -----CCCTATTATATTGGATTTGGTGTGATTAATCAAAATCAAC 903  
 QY 258 AlaLeuHisGluGlu-----ArgValGluArgProProMetPheProTyr 273  
 Db 904 ACCATTTTGAAGAAGATTGAATTTGTAGCCACTGAATCCCTTACCTTGCTCCCACTACA 963  
 QY 273 ----- 273  
 Db 964 ACTCCAATCGCTAATCAACTTGAAGATATTCCAATTTCTAAAGGTCATGAAGCTGCTTTA 1023  
 QY 273 ----- 273  
 Db 1024 ATCAACTTGCAAAAAATTTACCTTACCCTCAATCTATCAAAAGGAATATTGGAGA 1083  
 QY 274 -----PheLeuTyrThr 277  
 Db 1084 GTCTACTATGATGAATGAATCCATTGTATGAACAATTTAAAAACCAATACATATTATGCT 1143  
 QY 278 GlnSerTyrGluAspProGluProAspMetGluValMetGluIleAsnProLysAspThr 297  
 Db 1144 TTCACCTTGGGAAGATCTCTGCTGAAGATCATAACTTTTGAATTTTACCAGTGTATGATCT 1203  
 QY 298 ValLeuThrLeuThrSerGlyCysAsnAlaLeuAsnLeu-----LeuValGlnGly 315  
 Db 1204 GTTTGGCTATTACTTACCTGCTGTGATAATATTTTGATTTAGTTATGCTAGTTTCCCAACCA 1263  
 QY 316 AlaGlyGlnValValSerValAspCysAsnProAlaGlnSerAlaLeuLeuGluLeuLys 335  
 Db 1264 CCAAAAAAGATTCATGCTCTGTTGATCTTATCCATGTCAAAACCATTTATTAGATTGAAA 1323  
 QY 336 LysValAlaIleGlnGlnLeuGluPheGluAppValTyrGlnLeuPheGlyGluGlyVal 355  
 Db 1324 TTGGCTAGTTTGTAGATGCTCTTCTCAAGAAACAAATTTGGTCAATTTTGGTGAAGGTAAA 1383  
 QY 356 HisProArgIleGluGluLeuTyrGluLysLeuAlaProPheLeuSerGlnThrSer 375  
 Db 1384 ATCGAAAAATTCATGATCTTTTGTGATGATCTTTGGCCCGCACATGCTTCTTAATGCC 1443  
 QY 376 HisAsnPheTyrSerLysArg-----LeuTyrTyrPheGlnHisGlyLeuTyrGln 393  
 Db 1444 TTCCAATATGATGATGATGAAGACCTAAACCTTTTCTGGTAAAGGTCTTTAT----- 1497  
 QY 394 GlyGlyMetGlyLysLeuCysTyrValLeuGlnCys-----LeuAlaValValLeu 410  
 Db 1498 ---GATACCTGGGTTTCTAGATGGGCATTTAAGATTATCAAGATATGTTTTCAAAAGTTGC 1554  
 QY 411 GlyLeuGlyLysThrValLysArgLeuAlaAsnAlaProThrMetGluGluGlnArgArg 430  
 Db 1555 GGTGTTTAGTAAATACCTTGAAGAACCTTTGTGCTGCAACCCATCGAAGAACCAATTGAGA 1614  
 QY 431 LeuTyrAspSerAsnMetLeuIleHisPheValLysAsnGlyProLysProLeuValTyr 450  
 Db 1615 ATTTGAATGAACATTTG-----AAACCACT----- 1641  
 QY 451 LeuPheValLysPheVal---SerLeuValLeuPheAsnLysAlaValLeuTyrPheGly 469



```

Db 1642 TTATTAATCTGTGGTGGTTCATTACTTGTGGTAAATCCAAATGTTTATGCAAGCT 1701
Qy 470 GlyGlyValProGlyLysGlnTyrAlaLeuIleLysAlaAspGlyIleProIleGluAsn 489
Db 1702 TTGGGAGTTCAGCTAATCAAGCAGCTTAAATG-----GGACCATCTGTATTAA 1752
Qy 490 TyrIleAlaArgThrMetAspGlyValAlaGluAsnSerHisValArgLysGlnAsnTyr 509
Db 1753 TAGTGTGTGATCTTTGGACCCCATCATTAAGNGATCGATGATTTCAATGATTAATAT 1812
Qy 510 PheTyrTyrAsnCysLeuThrGlyLysPheLeuArgAspAsnCysProThrTyrLeuArg 529
Db 1813 TTCTACTATTATGATGATGGGAGATACACCAAAACAATGTCCAGATTATTAACT 1872
Qy 530 GluAlaAlaPheAlaThrLeuLys-----SerGlyValVal 541
Db 1873 ACAAAGGTTTCAACAGATTATCTAGCACTGTCTACTGCCAGCGGATCATCTCCAATT 1932
Qy 542 AspAsnLeuThrValSerThrAsnPhePheMetGly-----GluLeuLysAlaArg 558
Db 1933 GACAACTTAGAATCCACATGACACTTTAAATGAAGTGTGTGTAGATTAAAGAAAAA 1992
Qy 559 ThrTyrThrLysValIleLeuMetAspHisValAspTrpLeuAsp-----MetPro 575
Db 1993 TCRACTACTATTGCAATATCATGATCATATGATGTTTGCACCCCTAATGGTAGAGAT 2052
Qy 576 ValAlaAsnGluLeuAlaGluCysLeuAlaLysGlnValAlaProGlyIleValIle 595
Db 2053 GCTATTAAATGAATTAATCT---GCTTTGAAAAGAGATGCTTCCGCCAGGGGTAGATTA 2109
Qy 596 TrpArgSerAlaSerLeuSerProTyrAlaGluLeuIleGlnLysAlaGlyPhe--- 614
Db 2110 CTTAGATCAGCAAGTACAAGCCCTGTGTACTTGAACATTCAGAATCTGGGATCCAA 2169
Qy 615 AspValArgCysIleArgAlaAlaThrGlnGlyTyrMetAspArgValAsnMetTyrSer 634
Db 2170 GAAGAAGAAAAATGTTGTGTGCTCAACCTGGTTCAAGTATATAGACAGAGTTAATATGATGCC 2229
Qy 635 Ser 635
Db 2230 AAT 2232

```

## RESULT 4

```

US-10-620-914-49
; Sequence 49, Application US/10620914
; Publication No. US20040093639A1
; GENERAL INFORMATION:
; APPLICANT: Benning, Christoph
; APPLICANT: Riekhof, Wayne
; APPLICANT: Klug, Rouven
; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids
; FILE REFERENCE: MSU-07769
; CURRENT APPLICATION NUMBER: US/10/620,914
; CURRENT FILING DATE: 2003-07-16
; PRIOR APPLICATION NUMBER: 10/118,495
; PRIOR FILING DATE: 2002-04-08
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 49
; LENGTH: 2727
; TYPE: DNA
; ORGANISM: Neurospora crassa
US-10-620-914-49

```

```

Alignment Scores:
Pred. No.: 5,71e-94 Length: 2727
Score: 850.50 Matches: 239
Percent Similarity: 44.4% Conservative: 109
Best Local Similarity: 20.5% Mismatches: 240
Query Match: 24.6% Indels: 198
DB: 7 Gaps: 23

```

```

US-10-620-914-45 (1-648) x US-10-620-914-49 (1-2727)
Qy 19 LeuGluLysLeuLysLeuSerSerMetLysAspLeuThrValLeuArgHisMet--- 37
Db 449 TTGAGAGAGCTGGAGATG---GGCTTGAANAAGACAAGAGATGGAGACATGTGTAA 505
Qy 38 ---TrpPheGlySerLysLysGlyAspAspHisAlaAlaArgLeuGluSerPheTyrGly 56
Db 506 CCGTGGCCGGACAG-----GGA 523
Qy 57 ProGlnAlaAlaPheAlaAlaArgLeuAlaGluArgSerAsnLeuIleTrpValAsp 76
Db 524 CCG-----GGCAAGAGGAGAAACCG-ATATGGGTAGAT 555
Qy 77 LeuGlyGlyGlyThrGlyGluAsnValAspMetAlaAspTyrIleAspLeuAlaLys 96
Db 556 GTCGGTGGGGGCACAGGCTGGATATATCGAACCATGGCCAAAGTTTGTCAAGCTCTCGAA 615
Qy 97 ---PheLysSerIleTyrValValAspLeuCysHisSerLeuCysGluValAlaLysLys 115
Db 616 TTCTTCAAGACTGTTTTACCTAGTGGACTTTTCGCCGTCACTTTGTGAAGTGGCTAGGAAG 675
Qy 116 LysAlaLysAlaLysGlyTyrLysAsnValGlnValValGluAlaAspAlaCysGlnPhe 135
Db 676 CGGTTTGGCCAGGCTGGGGTGGGAGAAATGTGAGAGTTATCTGCACGGATGTCGCAAGTTT 735
Qy 136 Ala-----ProProGluGlyThr----- 141
Db 736 AGGCTTGAGGATTTATGAGGATGTTGAGAAAGAGAGTCTGGCTCTGGAGATTTCTCGCCT 795
Qy 142 -----AlaThrLeuIle 145
Db 796 TCCTTTGCGGTGGTGGGGGAGACGAAGCCGGACACATGCGGGAGCTGAGTTGCATC 855
Qy 146 ThrPheSerTyrSerLeuThrMetIleProPheHisAsnValIleAspGlnAlaCys 165
Db 856 ACCATGTCTTATAGCCTTTCGATGATGCCGGATTTCTTCGATTATCGATTTCGCTCGAG 915
Qy 166 SerTyrLeuSerGlnAspGlyLeuValGlyValAlaAspPheTyrValSerGlyLysTyr 185
Db 916 TCCTCTGTAGACCTCACGGCTTGATGCGCTGTGGACTTTTACGCCCGAGTGGAAAGTC 975
Qy 186 Asp-----LeuProLeuArgGlnMetProTyrSerArg 196
Db 976 GACTTCACATTCGCAACTACACGGGTGGTCTTATGAACGACACGTTGGCTATTTCGCG 1035
Qy 197 ArgPhePheTrpArgSerIlePheAspIleAspAsnIleAspIleGlyProGluArgArg 216
Db 1036 CGGAACCTTCTGGCGCTCGTGGTTGCGATGCTGACAGGGGTGTCTCTGAGCCAGCTCGTGA 1095
Qy 217 AlaTyrLeuGluGlnLysLeuGluArgValTrp-----GluGlnAsnThrGln 232
Db 1096 GATTAICTGAGTACAGTTCGGGACTGCTCAACCGTCAACGCCCGCAACACACTTGT 1155
Qy 233 GlySerIleProTyrValProTyrLeuArgAlaPro-----TyrTyrVal--- 247
Db 1156 GGAGCAATTCCTTACTACATCTGTTGGG-ATGCTCAGAAGAGCCCTTTTCTACGTGCGAG 1214
Qy 248 -----TrpIleGlyArgLeuPro----- 253
Db 1215 TCTACCACAGAAATTTGGAAACACATCGATGCTATTGTCGACAGAGTCCCCAAGATCATC 1274
Qy 253 ----- 253
Db 1275 ACCCGCTCTAGTGGGCAACATCTTCTCCTCAGCAACAAATGCGGTAGCCTTTGCGAGTCGG 1334
Qy 254 -----SerValGlyHisAlaLeu-----HisGluGluArgValGlu 265
Db 1335 CCGCACAGCGCGGAGATGCGCTCAAAAGGCGCTTCAATACGCGCATCGAGAACAATCTCGGC 1394
Qy 266 ArgPro-ProMet----- 269
Db 1395 CAACCTACCTCTCCCGTCCCTTCTTCTTACCAAAATCACCACTGGAGGATCTTACTAGACGA 1454

```

QY 270 -----PheProThrPheLeuThrGlnSerTrpGluAs 282  
Db 1455 TCAACTCCCGAAGCACACCCAGTTCAATGACGAGTACATCTACGCTTTTACCTGGGAAGA 1514  
QY 282 pProGluProAspMetGluValMetGluIleAsnProLysAspThrValLeuThrLeuTh 302  
Db 1515 CTGCGCGTGCAGAGAACTCTTAACTCGGCGCCGACGAGCTGCTCTACCATCAC 1574  
QY 302 rSerGlyGlyCysAsnAlaLeuAsnLeuValGlnGlyAlaGlyGlnValValSerVa 322  
Db 1575 CAGCGCGGCGACAACTTTCTTCTTACCTGATGACAGAGTCCCGCTCGGTGACCGCAT 1634  
QY 322 lAspCysAsnProAlaGlnSerAlaLeuLeuGluLeuLysValAlaIleGlnGluLe 342  
Db 1635 CGACCTAAACCCAGCCCAACACCTGCTTGAACCTCAAGTCCCTCTTTACGACTCT 1694  
QY 342 uGluPheGluAspValTrpGlnLeuPheGlyGluGlyValHisProArgIleGluLe 362  
Db 1695 GGATTACCCGACGCTCTGGAAGATCTTCGTGAGGGCAACACCCCGACTTTTCGCTCACT 1754  
QY 362 uTrpGluLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrpSerLysAr 382  
Db 1755 GCTCATCTCCAACTCTCCCTCACCTTCGCGCGCGGCTTCAATACCTAGCTGATCCAA 1814  
QY 382 gLeuTrpTrpPheGln-----HisGlyLeuTrpTrpGlnGlyGlyMetGlyLy 398  
Db 1815 TGGCACATATTTACGACCCCTCGGCGCGGCTCTATGATACCGGCGCTCCCGATA 1874  
QY 398 sLeuCysTrpValLeuGlnCysLeuAlaValValLeuGlyLeuGlyLysThrValLysAr 418  
Db 1875 CGTATACCGTTCTTCGCTGGATTTCCACACTCTTCTTCGCGCTCCGCGCTCGCTG 1934  
QY 418 gLeuAlaAsnAlaProThrMetGluGluGlnArgArgLeuTrpAspSerAsnMetLeu1 438  
Db 1935 ACTTCTCTTACTCCACCTCGAAGGCAAGCTTCATCTACACCAAGATT----- 1989  
QY 438 eHisPheValLysAsnGlyProLysProLeuValTrpLeuPheValLysPheValSer-- 457  
Db 1990 -----CGTCCCTGT-----CTGCTCAACCGCTTCGTCAACGG 2021  
QY 458 -LeuValLeuPheAsnLysAlaValLeuTrpPheGlyGlyGlyValProGlyLysGlnTy 477  
Db 2022 CTTGTCTCTCAGCTCCGACGCTTCTCTGCTGGCTTTTGGCGGTGCCCCAAGAAATCAAGT 2081  
QY 477 zAlaLeuLysAlaAsp----- 483  
Db 2082 GGCTATGATCGAAGCCGACTTACCACCGCGCTTCTATCTCTCTCCACACCCCGAGCAG 2141  
QY 484 -----GlyIleProIleGluAsnTrpIleAlaArgThrMetAspG1 497  
Db 2142 CAAAGAAAACCAAGCGCGCCGAGCAATCTCCACTACACACCTCCACCTTGATCC 2201  
QY 497 yValAlaGluAsnSerHisValArgLysGlnAsnTrpPheTrpAsnCysLeuThrG1 517  
Db 2202 CGTTCTCTCCACCTCCACCTTCGCTCGGCAACCTTACTACTCTGCTGTCTCTCTGGG 2261  
QY 517 yLysPheLeuArgAspAsnCysProThrTrpLeuArgGluAlaAlaPheAlaThrLeuLy 537  
Db 2262 ACAATACACGCGCAGTGCCATCCCGATTACCTTTCCCTCGCCGCCACCTCTATCTACTCAG 2321  
QY 537 sSer---GlyValValAspAsnLeuThrValSerThrAsnPhePheMetGluGluLeu-- 555  
Db 2322 CGTCTCTGGAGCCTTTGACGGCTTACGATCCACCGATGAATACAGGAGGTGTGGC 2381  
QY 556 -----LysAlaArgThrTrpThrLysValIleLeuMetAspHisValAspTrpLeuAs 573  
Db 2382 TAGTCTTCAGCCGGTACTTTGACAGTAGCGGTGGTGTGATGATGATGATGATGATGATG 2441  
QY 573 pMetPro-----ValAlaAsnGluLeuAlaGluCy 583  
Db 2442 TCCGCTCTTCGCTGAGGAGGAAAGGAAGGAGGCGCAAGCGGAGCAAGTGAGGAG 2501

QY 583 sLeuAlaLysGlnValAlaProGlyGlyIleValIleTrpArgSerAlaSerLeuSerPr 603  
Db 2502 GTTGAATCCGGCGTTCGAAGTGGTGGAAAGGTGTTGTTGAGGAGCGGAGTGGAGCC 2561  
QY 603 oProTrpAlaGluLeuLysAlaGlyPheAspValArgCysIle----- 619  
Db 2562 GTGGTATGTGAGGTTTGTGAGGAAGGTTTGGAGCAAGGTTGGAGCGGTGGTGTGCTGA 2621  
QY 620 ----ArgArgAlaThrGlnGlyTrpMetAspArgValAsnMetTrpSerSerPheTrpMe 638  
Db 2622 ATCCGGAAGGGGACAGGAGTGTATTGACAGGGTGAATATGATGCTAGTTGTTGGAT 2681  
QY 638 tAlaArgArg 641  
Db 2682 CTGGGAAG 2691  
RESULT 5  
US-10-620-914-48  
; Sequence 48, Application US/10620914  
; Publication No. US20040093639A1  
; GENERAL INFORMATION:  
; APPLICANT: Benning, Christoph  
; APPLICANT: Riekhof, Wayne  
; APPLICANT: Klug, Rouven  
; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids  
; FILE REFERENCE: MSU-07769  
; CURRENT APPLICATION NUMBER: US/10/620,914  
; CURRENT FILING DATE: 2003-07-16  
; PRIOR APPLICATION NUMBER: 10/118,495  
; PRIOR FILING DATE: 2002-04-08  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: Patent in version 3.2  
; SEQ ID NO 48  
; LENGTH: 3427  
; TYPE: DNA  
; ORGANISM: Neurospora crassa  
US-10-620-914-48  
Alignment Scores:  
Pred. No.: 5,11e-88 Length: 3427  
Score: 804.00 Matches: 238  
Percent Similarity: 43.2% Conservative: 112  
Best Local Similarity: 29.4% Mismatches: 256  
Query Match: 23.2% Indels: 205  
Gaps: 24  
US-10-620-914-45 (1-648) x US-10-620-914-48 (1-3427)  
QY 19 LeuGluLysLeuLysLeuSerSerMetLysAspAspLeuThrValLeuArgHisMet--- 37  
Db 776 TTGGAGGAGCTGGAGATG---GGCTTGAAGAAAGACAAAGAAATGGGAAGACATGTGTAA 832  
QY 38 ---TrpPheGlySerLysLysGlyAspAspHisAlaAlaArgLeuGluSerPheTrpGly 56  
Db 833 CGGTGCGCGGACAGGACCGGACCAAGGA-----GGAAACCATATGCG 877  
QY 57 ProGlnAlaAlaPheAlaAlaArgLeuAla-----GluArgSerAsn 71  
Db 878 TAGATGTACGCTCTCTCATCAGAACCTTGCCTTGAATGTCACACACGCTGACC 937  
QY 72 LeuIleTrpValAspLeuGlyGlyThrGlyGluAsnValAspMetMetAlaAspTrp 91  
Db 938 ATCGCTACAAAACAGGTGCGTGGGGGCACAGGCTGGAATATCGAAGCATGGCCAAATTT 997  
QY 92 IleAspLeuAlaLys---PheLysSerIleTrpValValAspLeuCysHisSerLeuCys 110  
Db 998 GTCACGCTCTGAATTTCTCAAGACTTTTACCTAGTGACTTTTCGCCGCTCACTTGT 1057  
QY 111 GluValAlaLysLysLysAlaLysAlaLysGlyTrpLysAsnValGlnValValGluAla 130  
Db 1058 GAAGTGGCTAGGAAGCGGTTTGCAGGCTGGGTGGAGATGTGAGAGTTATCTGCAGC 1117  
QY 131 AspAlaCysGlnPheAla-----ProProGluGlyThr----- 141



```

; APPLICANT: Klug, Rouven
; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine
; FILE REFERENCE: MSU-06897
; CURRENT APPLICATION NUMBER: US/10/118,495
; CURRENT FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: 60/283,812
; PRIOR FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1252
; TYPE: DNA
; ORGANISM: Rhodobacter sphaeroides
US-10-118-495-1

Alignment Scores:
Pred. No.:      7,19e-36      Length:      1252
Score:          379.50       Matches:     127
Percent Similarity: 43.1%    Conservative:   68
Best Local Similarity: 28.1% Mismatches:    182
Query Match:      11.0%     Indels:       75
DB:               5         Gaps:        20

US-10-620-914-45 (1-648) x US-10-118-495-1 (1-1252)

QY      231 ThrGlnGlySerIleProTyrValProTrpLeuArgAlaProTyrTyrValTrpIleGly 250
      ||||| :||||: :||||: :||||: |||||
Db      4  ACGCAGTTGGCCCTCACCCACTGCC-----GCCCGCGC-----GTTGC 45

QY      251 ArgLeuProSerValGlyHisAlaLeuHis-----GluGluArg 263
      :|||: :|||: :|||: :|||: :|||:
Db      46 CGC-----CAGATCGGCGCGCGCTGCACCGCACGTCGTCTTCACGCGCCGAAGACTG 99

QY      264 ValGluArgProMetPheProThrPhe-----LeuTyrThrGlnSerTrp 280
      :|||: :|||: :|||: :|||: :|||:
Db      100 ATGGAGCGG-----ATGTTCTCGGGCTCTTCACGGCCTGCTATCGCAGATCTGG 153

QY      281 GluAspProGluProAspMetGluValMetGluIleAsnProLysAspThrValLeuThr 300
      :|||: :|||: :|||: :|||: :|||:
Db      154 GAGGATCGCGCGGTGGACATGCGCGCCTCGCCATCGCCCCGGGACCGGCTGGTGCC 213

QY      301 LeuThrSerGlyGlyCysAsnAlaLeuAsnLeuValGlnGlyAlaGlyGlnValVal 320
      ::|||: :|||: :|||: :|||: :|||:
Db      214 ATCGCCTCGGGCGGTGCAACGTCTTTCTCATCTCACGCGAGGGCGCGCTGACTCTC 273

QY      321 SerValAspCysAsnProAlaGlnSerAlaLeuGluLeuLysLysValAlaIleGln 340
      :|||: :|||: :|||: :|||: :|||:
Db      274 GCGGTGGATCTCTGCGCGCCCATGTGCGCTGGGCGGCTGAAGCTCGCGCCGCGCG 333

QY      341 GlnLeu---GluPheGluAspValTrpGlnLeuPheGlyGluGlyValHisProArgIle 359
      :|||: :|||: :|||: :|||: :|||:
Db      334 ACGTGTCCCGACCATTGCGCGCTCTTCGATCTCTTCGTCGCGACACCTGCCCGCAAT 393

QY      360 GluGluLeuTyrGluLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrp 379
      :|||: :|||: :|||: :|||: :|||:
Db      394 GCGGCCCTCTACGACCGCCATATCGCGCCGCGCTCGACGCGCGAGCGCGCTACTGG 453

QY      380 -----SerLysArgLeuTrpTyrPheGlnHisGlyLeuTyrGln 393
      :|||: :|||: :|||: :|||: :|||:
Db      454 GAGGCGCGACGCCCTTCGCGCGCGCATCGACTGTTCAGCGCGCGCTTCTACCGGCAC 513

QY      394 GlyGlyMetGlyLysLeuCysTrpValLeuGlnCysAlaValValLeuGlyLeuGly 413
      :|||: :|||: :|||: :|||: :|||:
Db      514 GTGCGCTCGGCCGCTTCATCGGCGGCCCATACGCTCGG-----CGGGCGCGGGCG 567

QY      414 LysThrValLysArgLeuAlaAsnAlaProThrMetGluGlnArgArgLeuTrpAsp 433
      :|||: :|||: :|||: :|||: :|||:
Db      568 ACCGACCTCGGGGCTTCTCGACTGTCCGACATCGAGGCGCGAGCTTCTTCTCTAC 627

QY      434 SerAsnMetLeuIleHisPheValLysAsnGlyProLysProLeuValTrpLeuPheVal 453
      :|||: :|||: :|||: :|||: :|||:
Db      628 GCCCATATC-----GGGCGC-----CTCTTCGAG 651

```







	669	GACCTCGGTAAGGCCTCGCTG-----TTCCGGCTCGG	701
Db		:::	
Qy	471	yValProGlyLysGlnTyr-----AlaLeuileLysAlaAspGlyIleProileGluAs	489
		:::       :::	
Db	702	CATTCCGCGCGGCAGTAGCATTCCTGTGATCACCTCAGGCGACGCCACCATGCCAGCGT	761
Qy	489	nTyrlleAlaarg-----ThrMetAapGlyVvalalagLuasnSerHisValar	505
		:::	
Db	762	TCTGAAGCCCCGGCTGGAAAGCTCGCCTCGCAATTTCCCCTGGAANAAC-----	810
Qy	505	GlysGlnAsnTyPheTyrtYrAsnCysLeuThrGlyLysPheLeuArgAspAsn-----	523
		:::       :::	
Db	811	-----AATTATTTCCCTGGCAGGCTTTTGCCCGCGCTATCCAAATCCCGGTGAGGC	863
Qy	524	----CysProThrTyrLeuArgGluAlaAlaPheAlaThrLeuLysSerGlylValValas	542
		:::       :::	
Db	864	CGCCCTGCCCGCCTATCTGAAAAGCAGAACTACGAAACCATCCGC---GCGAATATCGA	920
Qy	542	pAsnLeuThrValSerThrAsnPhePheMetGlUGluLeu-----LysAlaargTh	559
		:::	
Db	921	CCGGTCGCCATCCACCATCCCAATCTGATCGAATCTCTCGCGCAAGACGCGGGCAC	980
Qy	559	rTyThrLysValleLeuMetAepHisValAspTrpLeuAspMetProValalalaenGI	579
		:::             :::	
Db	981	CGTCGATCGGTTTCCTCTGCTCGATGCGCAGGACTGGATGACCGATGACAGCTCAACGC	1040
Qy	579	uLeuAlaGluCyseuAlaLysGlnValAlaProGlyGlyIleValleltpargSerAl	599
		:::       :::	
Db	1041	GCTGTGTCGGAATCAGCGCACCGGCTCGCAGGGCGCCGGTCTCATCTTCGCGACCCG	1100
Qy	599	a-----SerIeuSerProProfTyralaGluLeuileGlnLysAlaGlyPheSep-V	616
		:::       :::	
Db	1101	CGCGAGCCCGACCTGTGTCGAGCGCCGCTCTCGACCTCGCTGCTCGACCACTGGGACTA	1160
Qy	616	alArgCyslleArqAcAlathrGlnGlyTyrrMetaspArgValAsnMerTyrrSerSerp	636
		:::       :::	
Db	1161	TCAGGACGAGGCGTCGCGCGAATTTCTCGGCACGC-GACCGTTGGCCCATCTATGGCGGCT	1219
Qy	636	hetYrMetalahArgArgLysGlyAla	644
		:::       :::	
Db	1220	TCCACCTCTATGTGAAGCGCAGCGCA	1245

```

RESULT 10
US-10-118-495-28
; Sequence 28, Application US/1018495
; Publication No. US20030074688A1
; GENERAL INFORMATION:
; APPLICANT: Benning, Christoph
; APPLICANT: Riekhof, Wayne
; APPLICANT: Klug, Rouven
; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids
; FILE REFERENCE: MSU-06897
; CURRENT APPLICATION NUMBER: US/10/118,495
; CURRENT FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: 60/283,812
; PRIOR FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28
; LENGTH: 1251
; TYPE: DNA
; ORGANISM: Agrobacterium tumefaciens
US-10-118-495-28

```

Alignment Scores:		
Pred. No.:	2,869-29	1251
Score:	376.50	1115
Percent Similarity:	47.4%	71
Conservative:	29.4%	162
Best Local Similarity:	29.4%	45
Query Match:	9.4%	14
DB:	5	



```
Qy 611 sAlaGlyPheAspValArgCysLeuArgAlaThrGlnGlyTyrMetAspArgValas 631
Db 1154 AGTGGGTCTATCTCGAAGAGCGCTCCAAACGA---ACTCAAC--GCCATGGACCGCTCGGC 1208

Qy 631 nMetTyrSerPheTyrMetAlaArgArg 641
Db 1209 CATTTATGGCGGCTTCCATATCTACCAGAGG 1239

RESULT 11
US-10-620-914-28
; Sequence 28, Application US/10620914
; Publication No. US20040093639A1
; GENERAL INFORMATION:
; APPLICANT: Benning, Christoph
; APPLICANT: Riekhof, Wayne
; APPLICANT: Klug, Rouven
; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids
; FILE REFERENCE: MSU-07769
; CURRENT APPLICATION NUMBER: US/10/620,914
; PRIOR FILING DATE: 2003-07-16
; PRIOR APPLICATION NUMBER: 10/118,495
; PRIOR FILING DATE: 2002-04-08
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 28
; LENGTH: 1251
; TYPE: DNA
; ORGANISM: Agrobacterium tumefaciens
US-10-620-914-28

Alignment Scores:
Pred. No.: 2,86e-29 Length: 1251
Score: 326.50 Matches: 115
Percent Similarity: 47.6% Conservative: 71
Best Local Similarity: 29.4% Mismatches: 162
Query Match: 9.4% Indels: 46
DB: 7 Gaps: 14

US-10-620-914-45 (1-648) x US-10-620-914-28 (1-1251)

Qy 275 LeuTyrThrGlnSerTrpGluAspProGluProAspMetGluValMetGluLeuLeuAsnPro 294
Db 130 GTCTATCCGAGATCTGGGAAGACCCCGAGATCGACATCGATGGAAGCGATGGAGCTTGGCGAA 189

Qy 295 LysAspThrValLeuThrLeuThrSerGlyGlyCysAsnAlaLeuAsnLeuValGln 314
Db 190 GGCACCGCATCTGACATCGGCTCGGCGGCTGCAACATGCTGGCTTATCTCTCGCGC 249

Qy 315 GlyAlaGlyGlnValValSerValAspCysAsnProAlaGlnSerAlaLeuLeuGluLeu 334
Db 250 AACCCGGCCAGCATCGATGTGTGGACCTCAACCCGCCACCATCGGCTGGAACAAGCTG 309

Qy 335 LysLysValAlaLeuGlnGlnLeu---GluPheGluAspValTrpGlnLeuPheGlyGlu 353
Db 310 AAGTCGCTCCCTTCGCCATCTCGCCGCCCATCAGGATGTGGTGGCCACTTTCGGCGCGC 369

Qy 354 GlyValHisProArgIleGluLeu-TyrGluLysLeuLeuAlaProPheLeuSerG1 373
Db 370 GCCGG-CACCCGGCAGCAACAGCTCGGTATGACCGGTTTCATCGCCGAGCATCTCGATGC 428

Qy 373 nThrHisAsnPheTrpSerLys-----ArgLeuTrpTyrPh 386
Db 429 CACGACCAAGGCATCTGGTTCGAAGCGCACCTTCCGGCGCGCTCGCATTTCCGTGT 488

Qy 386 eGlnHisGlyLeuTyrTrpGlnGlyMetGlyLysLeuCysTyrTrpValLeuGlnCysLe 406
Db 489 CGACAGGAACATCTACCGGACCGGCTGCTCGCGGTTTCATCGGCGCGGCCACATCAT 548

Qy 406 uAlaValValLeuGlyLeuGlyLysThrValLysArgLeuAlaAsnAlaProThrMetG1 426
Db 549 GGCCCGGCTCGACGGGTGAACTACCC-----GAAATGGCCAAGACCGCGGACGTGGA 602
```

```
Qy 426 uGluGlnArgArgLeuTrpAspSerAsnMetLeuLeuHisPheValLysAsnGlyProLy 446
Db 603 CGAACAGCGGCTTTTTCGACAGAGGTTCGCGCGCTTTTC-----GACAA 650

Qy 446 sProLeuVal---TrpLeuPheValLysPheValSerLeuValLeuPheAsnLysAlaVa 465
Db 651 GCCGGTGGTGGCTGGCTGAGTACGAGGCGCAAGAGCTCGCTT----- 690

Qy 465 lLeuTrpPheGlyGlyGlyValProGlyLysGlnTyr-----AlaLeuLeuLysAl 482
Db 691 -----TTTCGGCTTGGCATTCGCGCGCGCCAGTATGACGAGCTGGCAAGCTTTCAG 743

Qy 482 aAspGlyIleProIleGluAsnTyrIleAlaArgThrMetAspGlyValAlaGluAsnSe 502
Db 744 CGACGCGC---ACGGTTGCTCCTCGCTCAAGGAGCGGCTGGAAAGAGCTTGCCTGCAACTT 800

Qy 502 rHisValArgLysGlnAsnTyrPheTyrTrpAsnCysLeuThrGlyLysPheLeuArgAs 522
Db 801 CCGGCTC---AGGCACATTTATTCCTCGGCGGCTTTCGCGCGGCTTATCCGAGGCC 857

Qy 522 pAsn-----CysProThrTyrLeuArgGluAlaAlaPheAlaThrLeuLysSerG1 539
Db 858 GCATGAGGCTGCGCTGCGCTTATCTCAAGCGGAATATTACGAAAGATCCGCAACAA 917

Qy 539 yValValAspAsnLeuThrValSerThrAsnPhePheMetGluGluLeuLys----- 556
Db 918 CACCGCG---CGCGTCGCGTGCATCACGCCACCTATACCGAGCTGCTTTCGCGCAAGCC 974

Qy 557 -AlaArgThrTyrThrLysValIleLeuMetAspHisValAspTrpLeuAspMetProVa 576
Db 975 GGCAATGCGTGCAGCGCTATCTCTCGATGCGGAGCTGGATGATCGGATGTGCA 1034

Qy 576 lAlaAsnGluLeuAlaGluCysLeuAlaLysGlnValAlaProGlyGlyIleValIleTr 596
Db 1035 GCTCAACAGATTTATGTCGCGATCAGCGCATCGCGCATCGGGGCGCGCTCATCTT 1094

Qy 596 pArgSerAlaSerLeu-----SerProTyrAlaGluLeuLeuGlnLy 611
Db 1095 CCGCACCGCGCGCGCA-AAAGAGCGTTTTCGAGGCGGCTTTCGCGCGCATCCGCAACC 1153

Qy 611 sAlaGlyPheAspValArgCysIleArgAlaThrGlnGlyTyrMetAspArgValas 631
Db 1154 AGTGGGTCTATCTCGAAGAGCGCTCCAAACGA---ACTCAAC--GCCATGGACCGCTCGGC 1208

Qy 631 nMetTyrSerPheTyrMetAlaArgArg 641
Db 1209 CATTTATGGCGGCTTCCATATCTACCAGAGG 1239

RESULT 12
US-10-118-495-32
; Sequence 32, Application US/10118495
; Publication No. US20030074688A1
; GENERAL INFORMATION:
; APPLICANT: Benning, Christoph
; APPLICANT: Riekhof, Wayne
; APPLICANT: Klug, Rouven
; TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids
; FILE REFERENCE: MSU-06897
; CURRENT APPLICATION NUMBER: US/10/118,495
; CURRENT FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: 60/283,812
; PRIOR FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 32
; LENGTH: 1251
; TYPE: DNA
; ORGANISM: Sinorhizobium meliloti
US-10-118-495-32

Alignment Scores:
Pred. No.: 4,89e-25 Length: 1251
Score: 292.50 Matches: 106
```

Percent Similarity: 43.7% Conservative: 74  
Best Local Similarity: 25.7% Mismatches: 179  
Query Match: 8.4% Indels: 54  
DB: Gaps: 17

US-10-620-914-45 (1-648) x US-10-118-495-32 (1-1251)

Qy 263 ArgValGluArgProMetPhePro-----ProThrPhe 274  
Db 74 AAGCTCTCTCCCGCGGCTCTCCGAACGCTGTTCCGGGCTGCTCTTTTCCGGA-CTC 132

Qy 275 LeuTyrThrGlnSerTrpGluAspProGluProAspMetGluValMetGluLeuLeuLeuLeu 294  
Db 133 GTCTACCGGAGATCTGGGAGGACCGATGTCGACATGAAGCGATGCGATCCGTCCTCC 192

Qy 295 LysAspThrValLeuThrLeuSerGlyGlyCysAsnAlaLeuLeuLeuValGln 314  
Db 193 GGACATCGGATCGTGACGATCGGTTCCGGCGGCTGCAACATGCTGACCTATCTCTCCGCC 252

Qy 315 GlyAlaGlyGlnValSerValAspCysAsnProAlaGlnSerAlaLeuLeuLeuLeu 334  
Db 253 GAGCTTCCCGGATAGAGTGGTGGATCTCAACCCCATCATCGCGCTCAACCGGCTG 312

Qy 335 LysLysValAlaLeuGlnGlnLeu-----GluPheGluAspValTrpGlnLeuPheGly--- 352  
Db 313 AAGCTGTCTCCCTTCCGACCTGCGGAGCCACAGGACGCTGGTGGCGTCTCTCCCGCTC 372

Qy 353 GluGlyValHisProArgIleGluLeuTyrGluLysLysLeuAlaProPheLeuSer 372  
Db 373 GAAGTACCGCGCAGCAAT---GGCCAGGCTACGACGTTCTCTCCGCGGAGCTCGAT 429

Qy 373 GlnThrSerHisAsnPheTrpSer-----LysArgLeuTrpTyr 385  
Db 430 CCAGCAACCGCGCTATTGGAACGCGCGGAGATCTACCGCGCGCGCGCATCGGCGTC 489

Qy 386 PheGlnHisGlyLeuTyrTrpGlnGlyLeuGlyMetGlyLysLeuCysTrpValLeuGlnCys 405  
Db 490 TTCCGGCGCAACGTTATCTACCGGCTGCTGGCGGTTTCATTTCCGCGCAGCCATGCT 549

Qy 406 LeuAlaValValLeuGlyLeuGlyLysThrValLysArgLeuAlaLeuAlaProThrMet 425  
Db 550 CTGACAGCGGTGACGCGCATCAATCCG-----GAGATTTCTGTCAGGCGCGCTCCATG 603

Qy 426 GluGlnArgLeuTrpAspSerAsnMetLeuLeuHisPheValLysAsnGlyPro 445  
Db 604 CGCAGCAGCGGCGAGTCTTTCGACGACCAAGCTCGCTCCGCTCTTC-----GAG 651

Qy 446 LysProLeuVal---TrpLeuPheValSerValSerLeuValLeuPheAsnLysAla 464  
Db 652 CGTCCGTCATCCGTTGGATCACCAGCGGCAAGAGCTCCCTT-----693

Qy 465 ValLeuTrpPheGlyGlyValProGlyLysGlnTyrAlaLeuLysAlaAspGly 484  
Db 694 -----TTCCGCTCCGATCCCGCGGATCGAGTTCGACGAACTCGCGAGC-----738

Qy 485 IleProIleGluAsnTyrIleAlaArgThrMetAspGlyValAlaGluAsnSerHisVal 504  
Db 739 CTGACCGCGGAGAAATCCGTCGCGCGGTGCTGCGCAATCGCTCGGAAAGCTGACCTGT 798

Qy 505 Arg-----LysGlnAsnTyrPheTyrTrpAsnCysLeuThrGlyLysPheLeuArg 521  
Db 799 CATTTCCCTTCCCGGATTAATACTTTCGCTGGCAGCGCTTTCGACGCGGCTACCCCGG 858

Qy 522 AspAsn-----CysProThrTyrLeuArgGluAlaAlaPheAlaThrLeuLysSer 538  
Db 859 CCGACGAGGCGAGTGGCCACCTTATCTTCAGGATCGGATACGAAAGCGATTCGCGAC 918

Qy 539 GlyValValAspAsnLeuThrValSerThrAsnPhePheMetGluGluLeu-----555  
Db 919 AAT---GCGGAGCGCGTGGAGTCCACATCGGAGCTTACGAGCTTCTCGCGGCAAG 975

Qy 556 LysAlaArgThrTyrThrLysValIleLeuMetAspHisValAspTrpLeuAspMetPro 575  
Db 556 LysAlaArgThrTyrThrLysValIleLeuMetAspHisValAspTrpLeuAspMetPro 575

Db 976 CCGCGCGCTCAGTTCGACGCGCTACGTGCTCTCCGACGACGAGCTGGATGACCGACCAG 1035

Qy 576 ValAlaAsnGluLeuAlaGluCysLeuAlaLysGlnValAlaProGlyGlyIleValIle 595

Db 1036 CAGCTGAACGACCTCTGGACGGAGATCACCCGACCGCCGCGCGGCTGGTGGATC 1095

Qy 596 TrpArgSer-----AlaSerLeuSerProPro-----TyrAlaGluLeuLeuGln 610

Db 1096 TTCGCGACCGCGCGCGGAGGAGATCTCTCCGGGCGGCTCTCCACCGCTCTCTCGAT 1155

Qy 611 LysAlaGlyPheAspValArgCysIleLeuArgAlaThrGlnGlyTyrMetAspArgVal 630

Db 1156 CAGTGGTACTATGATCCGAGACTTCGATGAGCTC-----GGCGCTGAAGACCGGTG 1209

Qy 631 AsnMetTyrSerSerPheTyrMetAlaArgArgLys 642

Db 1210 GCGATCTATGCGGCTTCCACATCTACCGGAAGAAA 1245

RESULT 13

US-10-620-914-32

Sequence 32, Application US/10620914

Publication No. US20040093639A1

GENERAL INFORMATION:

APPLICANT: Benning, Christoph

APPLICANT: Riekhof, Wayne

APPLICANT: Klug, Rouven

TITLE OF INVENTION: Compositions and Methods for the Production of Betaine Lipids

FILE REFERENCE: MSU-07769

CURRENT APPLICATION NUMBER: US/10/620,914

CURRENT FILING DATE: 2003-07-16

PRIOR APPLICATION NUMBER: 10/118,495

PRIOR FILING DATE: 2002-04-08

NUMBER OF SEQ ID NOS: 52

SOFTWARE: Patent in version 3.2

SEQ ID NO 32

LENGTH: 1251

TYPE: DNA

ORGANISM: Sinorhizobium meliloti

US-10-620-914-32

Alignment Scores:

Pred. No.: 4,89e-25 Length: 1251

Score: 292.50 Matches: 106

Percent Similarity: 43.7% Conservative: 74

Best Local Similarity: 25.7% Mismatches: 179

Query Match: 8.4% Indels: 54

DB: Gaps: 17

US-10-620-914-45 (1-648) x US-10-620-914-32 (1-1251)

Qy 263 ArgValGluArgProMetPhePro-----ProThrPhe 274

Db 74 AAGCTCTCTCCCGCGGCTCTCCGAACGCTGTTCCGGGCTGCTCTTTTCCGGA-CTC 132

Qy 275 LeuTyrThrGlnSerTrpGluAspProGluProAspMetGluValMetGluLeuLeuLeuLeu 294

Db 133 GTCTACCGGAGATCTGGGAGGACCGATGTCGACATGAAGCGATGCGATCCGTCCTCC 192

Qy 295 LysAspThrValLeuThrLeuSerGlyGlyCysAsnAlaLeuLeuLeuValGln 314

Db 193 GGACATCGGATCGTGACGATCGGTTCCGGCGGCTGCAACATGCTGACCTATCTCTCCGCC 252

Qy 315 GlyAlaGlyGlnValSerValAspCysAsnProAlaGlnSerAlaLeuLeuLeuLeu 334

Db 253 GAGCTTCCCGGATAGAGTGGTGGATCTCAACCCCATCATCGCGCTCAACCGGCTG 312

Qy 335 LysLysValAlaLeuGlnGlnLeu-----GluPheGluAspValTrpGlnLeuPheGly--- 352

Db 313 AAGCTGTCTCCCTTCCGACCTGCGGAGCCACAGGACGCTGGTGGCGTCTCTCCCGCTC 372

Qy 353 GluGlyValHisProArgIleGluLeuTyrGluLysLysLeuAlaProPheLeuSer 372

Db 373 GAAGTACCGCGCAGCAAT---GGCCAGGCTACGACGTTCTCTCCGCGGAGCTCGAT 429

Qy 373 GlnThrSerHisAsnPheTrpSer-----LysArgLeuTrpTyr 385

Db 430 CCAGCAACCGCGCTATTGGAACGCGCGGAGATCTACCGCGCGCGCGCATCGGCGTC 489

Qy 386 PheGlnHisGlyLeuTyrTrpGlnGlyLeuGlyMetGlyLysLeuCysTrpValLeuGlnCys 405

Db 490 TTCCGGCGCAACGTTATCTACCGGCTGCTGGCGGTTTCATTTCCGCGCAGCCATGCT 549

Qy 406 LeuAlaValValLeuGlyLeuGlyLysThrValLysArgLeuAlaLeuAlaProThrMet 425

Db 550 CTGACAGCGGTGACGCGCATCAATCCG-----GAGATTTCTGTCAGGCGCGCTCCATG 603

Qy 426 GluGlnArgLeuTrpAspSerAsnMetLeuLeuHisPheValLysAsnGlyPro 445

Db 604 CGCAGCAGCGGCGAGTCTTTCGACGACCAAGCTCGCTCCGCTCTTC-----GAG 651

Qy 446 LysProLeuVal---TrpLeuPheValSerValSerLeuValLeuPheAsnLysAla 464

Db 652 CGTCCGTCATCCGTTGGATCACCAGCGGCAAGAGCTCCCTT-----693

Qy 465 ValLeuTrpPheGlyGlyValProGlyLysGlnTyrAlaLeuLysAlaAspGly 484

Db 694 -----TTCCGCTCCGATCCCGCGGATCGAGTTCGACGAACTCGCGAGC-----738

Qy 485 IleProIleGluAsnTyrIleAlaArgThrMetAspGlyValAlaGluAsnSerHisVal 504

Db 739 CTGACCGCGGAGAAATCCGTCGCGCGGTGCTGCGCAATCGCTCGGAAAGCTGACCTGT 798

Qy 505 Arg-----LysGlnAsnTyrPheTyrTrpAsnCysLeuThrGlyLysPheLeuArg 521

Db 799 CATTTCCCTTCCCGGATTAATACTTTCGCTGGCAGCGCTTTCGACGCGGCTACCCCGG 858

Qy 522 AspAsn-----CysProThrTyrLeuArgGluAlaAlaPheAlaThrLeuLysSer 538

Db 859 CCGACGAGGCGAGTGGCCACCTTATCTTCAGGATCGGATACGAAAGCGATTCGCGAC 918

Qy 539 GlyValValAspAsnLeuThrValSerThrAsnPhePheMetGluGluLeu-----555

Db 919 AAT---GCGGAGCGCGTGGAGTCCACATCGGAGCTTACGAGCTTCTCGCGGCAAG 975

Qy 556 LysAlaArgThrTyrThrLysValIleLeuMetAspHisValAspTrpLeuAspMetPro 575

Db 556 LysAlaArgThrTyrThrLysValIleLeuMetAspHisValAspTrpLeuAspMetPro 575

373 GlnThrSerHisAsnPhetTrpSer-----LysArgLeuTrpTyr 385  
386 PheGlnHisGlyLeuTyrGlnGlyMetGlyLysLeuCysTrpValLeuGlnCys 405  
490 TTCGGCGCAGCGGCTTATCGACCGGCTCTGTCGCTTCATTTCCCGCAGCCATGCT 549  
406 LeuAlaValValLeuGlyLeuGlyLysThrValLysArgLeuAlaAsnAlaProThrMet 425  
550 CTGCGACGGCTGACGCGATCATCCG-----GAAGATTTCGTCAGGGCGCTCCATG 603  
426 GluGlnGlnArgLeuTrpAspSerAsnMetLeuHisPheValLysAsnGlyPro 445  
604 CGCGAGCAGCGGCGGCTTCTTCGACGACAGCTCGCTCCGCTCTTC-----GAG 651  
446 LysProLeuVal---TrpLeuPheValLysPheValSerLeuValLeuPheAsnLysAla 464  
652 CGTCCGGTCTATCGCTTGGATCCAGCCGCAAGAGCTCCCTT----- 693  
465 ValLeuTrpPheGlyGlyGlyValProGlyLysGlnTyrAlaLeuLeuLysAlaAspGly 484  
694 -----TTCGGCTCGCATCTCCCGCAGAGGTTCCAGCACTCGCGAGC----- 738  
485 IleProIleGluAsnTyrIleAlaArgThrMetAspGlyValAlaGluAsnSerHisVal 504  
739 CTGAGCGCGGAGAAATCCGTCGCGCGTCTGCGCAATCGCTGGAAAGAGCTGACCTGT 798  
505 Arg-----LysGlnAsnTyrPheTyrTyrAsnCysLeuThrGlyLysPheLeuArg 521  
799 CATTTCCCTTGGCGGATACCTACTTCTCCCTCGCAGCGCTTTGCACGCGCTACCCGCGG 858  
522 AspAsn-----CysProThrTyrLeuArgGluAlaAlaPheAlaThrLeuLysSer 538  
859 CCGAGCAGCGGAGGTTGCGACCTTATCTTCAGCATCGCATACGAGCGATTCGGAC 918  
539 GlyValValAspAsnLeuThrValSerThrAsnPheMetGluGluLeu----- 555  
919 AAT---CGCGAGCGCGTCGAGTCCACCATCGCAGCTTTCAGCGGCTTCTCGCGCGCAAG 975  
556 LysAlaArgThrTyrThrLysValIleLeuMetAspHisValAspTrpLeuAspMetPro 575  
976 CCGCGCGCTTCAGTCGACCGCTAGTCTCTCGCAGCAGCAGCTGGATGACCGACCCAG 1035  
576 ValAlaAsnGluLeuAlaGluCysLeuAlaLysGlnValAlaProGlyGlyIleValIle 595  
1036 CAGCTGACGACCTCTGGACGAGATCACCGCAGCGCGCGCGCGCGCTCGTGCATC 1095  
596 TrpArgSer-----AlaSerLeuSerProPro-----TyrAlaGluLeuIleGln 610  
1096 TTCGACGCGCGCGCGAGCGAGCATCTCGCGGGCGCTCTCCACCCCTCTCCGAT 1155  
611 LysAlaGlyPheAspValArgCysIleArgArgAlaThrGlnGlyTyrMetAspArgVal 630  
1156 CAGTGTACTATGATGATCGGAGACTTCGATGAGGCTC-----GGCGCTGAAGACCGCTCG 1209  
631 AsnMetTyrSerSerPheTyrMetAlaArgArgLys 642  
1210 GCGATCTATGGCGCTTCCATCATCTACCGGAAGAAA 1245

## RESULT 14

US-10-085-959-12  
; Sequence 12, Application US/10085959  
; Publication No. US20030165870A1  
; GENERAL INFORMATION:  
; APPLICANT: Blattner, Frederick R.  
; APPLICANT: Welch, Rodney A.  
; APPLICANT: Burland, Valerie D.  
; TITLE OF INVENTION: No. US20030165870A1el Sequence of E. Coli CFT073  
; FILE REFERENCE: 960296.97648  
; CURRENT APPLICATION NUMBER: US/10/085,959  
; CURRENT FILING DATE: 2002-03-01

; PRIOR APPLICATION NUMBER: 60/242,412  
; PRIOR FILING DATE: 2000-10-19  
; NUMBER OF SEQ ID NOS: 255  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 12  
; LENGTH: 4246  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
US-10-085-959-12

## Alignment Scores:

Prod. No.: 0.00109 Length: 4246  
Score: 128.00 Matches: 139  
Percent Similarity: 34.1% Conservative: 108  
Best Local Similarity: 19.2% Mismatches: 207  
Query Match: 3.7% Indels: 272  
DB: 6 Gaps: 40

US-10-620-914-45 (1-648) x US-10-085-959-12 (1-4246)

QY 24 LeuSerSerMetLysAspLeuThrValLeuArgHisMetTrp-PheGlySerLysTly 43  
DB 2393 ATTGAACAGCGTAGGGAGGATATCGTATCTCTGCGCAACTGCAGTTTCTGCAACAAAC 2452  
QY 43 sGlyAspAspHisAlaAlaArgLeuGluSerPheTyrGlyPro-----G1 58  
DB 2453 CGGTGAGTCTCACTCA-----CCATACTTTCATTCCTGACGCGCAGCTGAAC 2500  
QY 58 nAlaAlaAlaPheAlaAlaArgLeuAlaGluArgSerAsnLeuLeuTrp----- 74  
DB 2501 AGGGCTGGCGTTTTCATTG---TGCACAGAAACCCCGAGCTAGGCTGGGGTTCGGAA 2557  
QY 75 -----ValAspLeu----- 77  
DB 2558 AGCTTTCAGCTTTCAGCCAGTTATTAAACCCCTTTTGTATTGTTTAAACACCTTCGGT 2617  
QY 78 -----GlyGlyGlyThrGlyGluAsnValAs 86  
DB 2618 CTGGCAACTGCAAGTGTCAAAACAAGAAATCAAAAGGGGTCTCCCAATGGGGAACGAAAGA 2677  
QY 86 pMetMetAlaAspTyrIleAspLeuAlaLysPheLysSerIleTyrValValAspLeuCy 106  
DB 2678 GCTTACGCGACACCCGATGGAACCTG-----TAAATATC 2710  
QY 106 shisSerLeuCysGlu-----ValAlaLysLysLysAla-LysAlaL 120  
DB 2711 ACATAGTTTTCGCCAAATAACCGAAGACAGGTGTTCTACAGAGAGAGCGTAGACAA 2770  
QY 120 yGly-----TrpLysAsnValGlnValValGluAlaAspA 132  
DB 2771 TAGGCAGTATTTTGAGAAAGCTGTGTGAGTGGAAAAGCGTACGATTCGTGAAGACTGAAT 2830  
QY 132 laCysGlnPheAlaProProGluGlyThrAlaThrLeuIleThrPheSerTyrSerLeuT 152  
DB 2831 GCTGT-----GCAGATCATATCCATATGCTTGT 2857  
QY 152 hrMetIleProProPheHisAsnValIleAspGlnAlaCysSerTyrLeuSerGlnAspG 172  
DB 2858 TGGAGATCCGCC-----AAATGAGCGGTATCCG 2887  
QY 172 lylLeuValGlyValAlaAspPheTyrValSerGlyLysTyrAspLeuProLeuArgGln- 191  
DB 2888 GGTATTATGGA-----TATCTGAAGGGGAAAGACGAGCTGTGATCTTTAGAGC 2935  
QY 192 -----MetProTrpSerArgArgPhePheTrpArgSerIlePheAspI 206  
DB 2936 AGTTTGGTATTGAAATTCAAATACAGAAACAGGAGGTCTGTCGACAGGGGTACTACG 2995  
QY 206 leAspAsnIleAspIleGlyProGluArgArgAlaTyrLeuGluGlnLysLeuGluArgV 226  
DB 2996 TCGATACATGGGT-----AAGAACACCGCGAAGATACAGGATTACATAAGACCC 3046  
QY 226 alTrpGluGlnAsnThrGlnGlySerIleProTyrValProTrpLeuArgAlaProTyr- 245

D	b		3947	- - - - - CTTCAGGGAAGAANTGCTATGAACATGAGATGGTTCATTATAAA	3991
Q	y		526	ThrTyrLeuArgGluaAlaPheAlaThrLeu- --LysSerGlyValValAspAsnLeu	544
D	b		3992	TACTTTAAGAGGGCTCGGCACATCACGTTCAGGAGCCAGAACCTGGTTTGACAGACTG	4051
Q	y		545	ThrValserThrAsnPheMetGluGluLeuLeuYsaIalaArgThrTyrThrLysValile	564
D	b		4052	- - - - - AATGATATTATGCCAAGAACA-	4072
Q	y		565	LeuMetAspHisValAspTrpLeuAspMetProValAlaAsnGluLeuAlaGluCys--	583
D	b		4073	CTATCTTGACATGCTGGCAGTATAGCTTTGCCAGTAGCTCAAAAGCTCGGAATTTGCCG	4132
Q	y		584	- - - - - LeuAlaLysGlnValala	594
D	b		4133	ATAAGATATGATCGTGGGGCATATAACC GGTTTCAGATGGCATACC CGTGGCATACT-	4191
Q	y		595	IleTrpArg 597	
D	b		4192	GTCGTGGAAG 4200	
 RESULT 15 US-10-684-141-57					
; Sequence 57, Application US/10684141					
; Publication No. US20050003536A1					
; GENERAL INFORMATION:					
; APPLICANT: Futusawa, Mitsuuru					
; TITLE OF INVENTION: METHOD AND SYSTEM FOR RAPIDLY CONFERRING					
; FILE REFERENCE: 690116.401					
; CURRENT APPLICATION NUMBER: US/10/684,141					
; CURRENT FILING DATE: 2003-10-10					
; NUMBER OF SEQ ID NOS: 66					
; SOFTWARE: FastSeq for Windows Version 4.0					
; SEQ ID NO 57					
; LENGTH: 7119					
; TYPE: DNA					
; ORGANISM: Mus musculus					
US-10-684-141-57					
 Alignment Scores: Pred. No.: 0.0106 Length: 7119 Score: 123.00 Matches: 144 Percent Similarity: 33.7% Conservative: 83 Best Local Similarity: 21.4% Mismatches: 214 Query Match: 3.6% Indels: 234 DB: 8 Gaps: 36					
 US-10-620-914-45 (1-648) x US-10-684-141-57 (1-7119)					
Q	y		107	HissertLeuCysGluValAlaLysLyseLysAlaLysAlaLys---GlyTrpLysAsnVal	125
D	b		1307	CATAATCTC-----AAGGCAGCTGCCAAGGCCAAACCTTGCTATGACCCCTGTA	1354
Q	y		126	GlnValValGluAlaAspAlaCysGlnPheAlaProProGluGlyThrAlaThrLeuile	145
D	b		1355	GAGCTGACCCTGAGGACATGTGCTGATGCCC---ACTGAAcAGCCCCAGACTCTGCCC	1411
Q	y		146	ThrpHeserTyrrgr	
D	b		1412	ACTTACTCAGTGTGAGTGTGCTGCTGCTGCTTACTTACTCTGATGATGAATACTGTCACCCC	1471
Q	y		150	- - - - -	150
D	b		1472	TTCATATTGCCCTGTGCAcCATTTATTC CATGGAACCTGATAGGTGTCGGGAAGGGC	1531
Q	y		151	- - - - - LeuThrMetIleProPhetis---AsnValille---	161
D	b		1532	TCCGGGACACTGTGTGAAGCCTTGCTGATGTGTGCAAGCTTTCCATGCGCAACATTATCTTC	1591
Q	y		162	- - - - - AspGlnAlaCyssertyrLeusserGlnaspGlyLeuValGlyValala	177



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: March 14, 2006, 06:44:17 ; Search time 1226 Seconds  
(without alignments)  
1219.517 Million cell updates/sec

Title: US-10-620-914-45

Perfect score: 3463

Sequence: 1 MSGRDRPASVTKKNFSL.....RVNYSFVARRKGAKN 648

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 7673375 seqs, 1153648444 residues

Total number of hits satisfying chosen parameters: 15346750

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-O=/abes/ABSSWEB spool/US10620914/runat\_13032006\_102052\_6964/app\_query.fasta.1  
-DB=Published Applications NA.New -QWMT=fastap -SUFFIX=p2n.rnpbn -MINMATCH=0.1  
-LOOPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-THR\_MIN=0 -ALIGN=15 -DOCLIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-MAXLEN=2000000000 -HOST=abes05h  
-USER=US10620914 @CCN 1.1.431 @runat\_13032006\_102052\_6964 -NCPU=6 -ICPU=3  
-NO WMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA.New:

1: /cgn2\_6/ptodata/2/pubpna/US08 NEW PUB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/US06 NEW PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US07 NEW PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/PCT NEW PUB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US09 NEW PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/US10 NEW PUB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US11 NEW PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US10 NEW PUB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US11 NEW PUB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US11 NEW PUB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US11 NEW PUB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US11 NEW PUB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US60 NEW PUB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	118.5	3.4	1155	12 US-11-234-786-373	Sequence 373, App
2	118.5	3.4	2000	12 US-11-234-786-374	Sequence 374, App
3	118.5	3.4	2040	12 US-11-234-786-375	Sequence 375, App
4	115.5	3.3	1512	12 US-11-234-786-368	Sequence 368, App

5	109.5	3.2	1853	12 US-11-234-786-369	Sequence 369, App
6	107.5	3.1	1879	12 US-11-234-786-531	Sequence 531, App
7	107.5	3.1	1852	12 US-11-234-786-530	Sequence 530, App
8	104	3.0	3006	7 US-10-932-182A-4215	Sequence 4215, App
9	104	3.0	3006	7 US-10-932-182A-4215	Sequence 4215, App
10	102	2.9	1857	7 US-10-932-182A-2428	Sequence 2428, App
11	102	2.9	1857	7 US-10-932-182A-2428	Sequence 2428, App
12	102	2.9	3042	12 US-11-124-367A-21	Sequence 21, Appl
13	101.5	2.9	7437	7 US-10-932-182A-1448	Sequence 1448, App
14	101.5	2.9	7437	7 US-10-932-182A-1448	Sequence 1448, App
15	101	2.9	3063	12 US-11-169-041-29	Sequence 29, Appl
16	99.5	2.9	2530	8 US-10-821-234-276	Sequence 276, App
17	99	2.9	2299	9 US-11-072-512-652	Sequence 652, App
c	18	2.9	1457619	12 US-11-098-686-8739	Sequence 8739, App
19	98	2.8	26214	7 US-10-330-773-222	Sequence 222, App
20	97.5	2.8	2315	9 US-11-072-512-1818	Sequence 1818, App
21	97	2.8	163162	12 US-11-121-086-66	Sequence 66, Appl
22	96.5	2.8	1851	12 US-11-234-786-366	Sequence 366, App
23	96.5	2.8	2184	12 US-11-234-786-370	Sequence 370, App
24	96.5	2.8	2634	12 US-11-136-527-1984	Sequence 1984, App
25	96	2.8	978	8 US-10-467-657-7223	Sequence 7223, App
26	96	2.8	2919	8 US-10-821-234-735	Sequence 735, App
27	96	2.8	3985	12 US-11-096-568A-2579	Sequence 2579, App
28	94	2.7	2368	9 US-11-000-688-623	Sequence 623, App
29	93.5	2.7	2722	12 US-11-055-877-7	Sequence 7, Appl
30	93.5	2.7	3835	8 US-10-055-877-9	Sequence 9, Appl
31	93.5	2.7	3985	8 US-10-055-877-9	Sequence 416, App
32	93.5	2.7	6072	8 US-10-995-561-416	Sequence 416, App
33	93.5	2.7	6113	8 US-10-995-561-415	Sequence 415, App
34	93.5	2.7	6601	8 US-10-995-561-414	Sequence 414, App
35	93.5	2.7	6659	8 US-10-995-561-417	Sequence 417, App
c	36	2.7	13806	8 US-10-995-561-13250	Sequence 13250, A
37	93.5	2.7	22471	8 US-10-995-561-13335	Sequence 13335, A
38	93.5	2.7	195998	8 US-10-995-561-13489	Sequence 13489, A
39	93	2.7	548	12 US-09-925-065A-129877	Sequence 129877, App
40	93	2.7	1254	12 US-11-098-686-9985	Sequence 9985, App
c	41	2.7	406	9 US-11-072-512-973	Sequence 973, App
42	92	2.7	1037	9 US-11-096-568A-1364	Sequence 1264, App
43	92	2.7	2250	7 US-10-932-182A-3450	Sequence 3450, App
44	92	2.7	2250	7 US-10-932-182A-3450	Sequence 3450, App
45	92	2.7	4092	7 US-10-932-182A-3000	Sequence 3000, App

#### ALIGNMENTS

RESULT 1

US-11-234-786-373  
; Sequence 373, Application US/11234786

; Publication No. US20060024301A1

; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun

; APPLICANT: Dillon, Davin C.

; APPLICANT: Mitcham, Jennifer L.

; APPLICANT: Harlocker, Susan L.

; APPLICANT: Jiang, Yuqiu

; APPLICANT: Reed, Steven G.

; APPLICANT: Kalos, Michael D.

; APPLICANT: Fanger, Gary R.

; APPLICANT: Retter, Marc W.

; APPLICANT: Stolk, John A.

; APPLICANT: Day, Craig H.

; APPLICANT: Vedvick, Thomas S.

; APPLICANT: Carter, Darrick

; APPLICANT: Li, Samuel X.

; APPLICANT: Wang, Aijun

; APPLICANT: Skeiky, Yasir A.

; TITLE OF INVENTION: PROSTATE-SPECIFIC POLYPEPTIDES AND FUSION

; FILE OF INVENTION: POLYPEPTIDES THEREOF

; FILE REFERENCE: 210121.427C31

; CURRENT APPLICATION NUMBER: US/11/234,786

; CURRENT FILING DATE: 2005-09-23

; PRIOR APPLICATION NUMBER: US 09/568,857

; PRIOR FILING DATE: 2000-05-09



```
; PRIOR APPLICATION NUMBER: US 09/536,857
; PRIOR FILING DATE: 2000-05-27
; PRIOR APPLICATION NUMBER: US 09/483,672
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: US 09/439,313
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: US 09/352,616
; PRIOR FILING DATE: 1999-07-13
; PRIOR APPLICATION NUMBER: US 09/288,946
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: US 09/232,149
; PRIOR FILING DATE: 1999-01-15
; PRIOR APPLICATION NUMBER: US 09/159,812
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: US 09/115,453
; PRIOR FILING DATE: 1998-07-14
; PRIOR APPLICATION NUMBER: US 09/030,607
; PRIOR FILING DATE: 1998-02-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 701
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 373
; LENGTH: 1155
; TYPE: DNA
; ORGANISM: Homo sapien
US-11-234-786-373

Alignment Scores:
Pred. No.: 0.0507 Length: 1155
Score: 118.50 Matches: 66
Percent Similarity: 37.4% Conservative: 19
Best Local Similarity: 29.1% Mismatches: 81
Query Match: 3.4% Indels: 61
DB: 12 Gaps: 11

US-10-620-914-45 (1-648) x US-11-234-786-373 (1-1155)

QY 187 LeuProLeuArgGlnMetProTrp-----SerArgArgPhePheTrpArgSerIlePhe 204
Db 204 CTTCCCTGCTGCGGGGAGTGGCAGACAGCAAGTGGGGCGCTTCGGAGA-----254

QY 205 AspIleAspAsnIleAspIleGlyProGluArgAlaTyrLeuGluGlnLysLeuGlu 224
Db 255 -----CCACGACGACTCTGCTAT-----GAA 275

QY 225 ArgValTrpGluGlnAsnThrGlnGlySerIleProTyrValProTrpLeuArgAlaPro 244
Db 276 GACACTCAGGAACAAGATGGGCAAGTGGTGTGCTGCCACTGCTCCCTGCTGCGAGGGGAG 335

QY 245 TyrTyrValTrpIleGlyArgLeuProSerValGly---HisAlaLeuHisGluGluArg 263
Db 336 CGGCAGAGCAAGTGGGGCGCTTGGGAGACTGAGATGACAGTGCCTTCATGGAGCCAG 395

QY 264 ValGluArgProMetPheProProThrPheLeuTyrThrGlnSerTrp-GluAsp-- 282
Db 396 GTACCAGCTCG-----TGAGAGAGATCT 419

QY 283 -----ProGluProAspMetGluValme 290
Db 420 GGCAAGCTCCACAGAGCTGCTGGTGGGTAAAGTCCCGAAGAGGATCTCATGCGCAT 479

QY 290 t-----GluIleAsnProLysAsp-----ThrValLeuThrLeuTh 302
Db 480 GCTCAGGGACACTGACGTGACGAGCAAGAGGAGCAAGCAAGCAAGAGGAGGAGTCTCTACATCTGGC 539

QY 302 rSer-----GlyGlyCysAsnAlaLeuAsnLeuValGlnGlyAlaGlyGlnValva 320
Db 540 CTCTGCCAATGGGAATTTCAGAGAGTAGTAAAACTCTCTGTCGACAGACGAGATGCAACTAA 599

QY 320 lSerValAspCysAsnProAlaGlnSerAlaLeuLeuLeuLysLysValAlaIleGl 340
Db 600 TGTCTTGTGAC---AACAAAGAGGAGGACGAGTCTGTATAAG-----GCCGTACA 644
```

```
QY 340 nGlnLeuGluPheGluAspValTrpGlnLeuPheGlyGluGlyValHisProArgIleGl 360
Db 645 ATGCCAGGAGAGATGATGCGTTAATGTTGCTGGAAACATGGCACTGATCCAAATATCC 704

QY 360 uGluLeuTyr-GluLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrpS 380
Db 705 AGATGAGTATGGAATACCACTCTGCACCTACGCTATCTATATGAAGATAAATTAATGGC 764

QY 380 erIysArgLeuTrpTyr 385
Db 765 CAAAGCACTGCTCTTAT 781

RESULT 2
US-11-234-786-374
; Sequence 374, Application US/11234786
; Publication No. US20060024301A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.
; TITLE OF INVENTION: PROSTATE-SPECIFIC POLYPEPTIDES AND FUSION
; TITLE OF INVENTION: POLYPEPTIDES THEREOF
; FILE REFERENCE: 210121.427C31
; CURRENT APPLICATION NUMBER: US/11/234,786
; PRIOR FILING DATE: 2005-09-23
; PRIOR APPLICATION NUMBER: US 09/568,857
; PRIOR FILING DATE: 2000-05-09
; PRIOR APPLICATION NUMBER: US 09/536,857
; PRIOR FILING DATE: 2000-05-27
; PRIOR APPLICATION NUMBER: US 09/483,672
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: US 09/439,313
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: US 09/352,616
; PRIOR FILING DATE: 1999-07-13
; PRIOR APPLICATION NUMBER: US 09/288,946
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: US 09/232,149
; PRIOR FILING DATE: 1999-01-15
; PRIOR APPLICATION NUMBER: US 09/159,812
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: US 09/115,453
; PRIOR FILING DATE: 1998-07-14
; PRIOR APPLICATION NUMBER: US 09/030,607
; PRIOR FILING DATE: 1998-02-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 701
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 374
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Homo sapien
US-11-234-786-374

Alignment Scores:
Pred. No.: 0.105 Length: 2000
Score: 118.50 Matches: 66
Percent Similarity: 37.4% Conservative: 19
Best Local Similarity: 29.1% Mismatches: 81
Query Match: 3.4% Indels: 61
```



```
Db 540 CTCTGCCAATGGGAATTCAGAACTAGTAAACTCTCTGTCGACAGCATGTCAACTTAA 599
QY 320 lSerValAspCysAsnProAlaGlnSerAlaLeuLeuGluLeuLysValAlaIleG1 340
Db 600 TGTCTCTTGAC--AACAAAAGAGGACAGCTCTGTATAAAG-----GCCGTACA 644
QY 340 nGlnLeuGluPheGluAspValTrpGlnLeuPheGlyGluGlyValHisProArgIleG1 360
Db 645 ATCCAGGAAGATGAATGTCGTTAATGTTGCTGGAACTGGCAGCATGATCCAATATTC 704
QY 360 uGluLeuTyr-GluLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrpS 380
Db 705 AGATGATATGGAATACCACTCTGCACACTCTGCACACTATCTATATAATGAAGATAATATG 764
QY 380 erLysArgLeuTrpTyr 385
Db 765 CAAAGCACTGCTCTTAT 781
```

## RESULT 4

```
US-11-234-786-368
; Sequence 368, Application US/11234786
; Publication No. US20060024301A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yugu
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.
; TITLE OF INVENTION: PROSTATE-SPECIFIC POLYPEPTIDES AND FUSION
; FILE REFERENCE: 210121.427C31
; CURRENT APPLICATION NUMBER: US/11/234,786
; PRIOR FILING DATE: 2005-09-23
; PRIOR FILING DATE: 2000-05-09
; PRIOR FILING DATE: 2000-05-09
; PRIOR FILING DATE: 2000-05-27
; PRIOR FILING DATE: 2000-01-14
; PRIOR FILING DATE: 1999-07-13
; PRIOR FILING DATE: 1999-04-09
; PRIOR FILING DATE: 1999-01-15
; PRIOR FILING DATE: 1999-11-12
; PRIOR FILING DATE: 1998-09-23
; PRIOR FILING DATE: 1998-07-14
; PRIOR FILING DATE: 1998-02-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 701
; SOFTWARE: Fast-Seq for Windows Version 3.0
; SEQ ID NO 368
; LENGTH: 1512
; TYPE: DNA
; ORGANISM: Homo sapien
US-11-234-786-368
```

```
Alignment Scores:
Pred. No.: 0.149 Length: 1512
Score: 115.50 Matches: 65
Percent Similarity: 36.6% Conservative: 18
Best Local Similarity: 28.6% Mismatches: 83
Query Match: 3.3% Indels: 61
DB: 12 Gaps: 11

US-10-620-914-45 (1-648) x US-11-234-786-368 (1-1512)
QY 187 LeuProLeuArgGlnMetProTrp-----SerArgArgPhePheTrpArgSerIlePhe 204
Db 426 CTTCCCTCTGTCAGGGGAGTGGCAAGACGTCGCGCTCTTGGAGA----- 476
QY 205 AspIleAspAsnIleAspIleGlyProGluArgArgAlaTyrLeuGluGlnLysLeuGlu 224
Db 477 -----CCAGCAGCATCTGCTAT-----GAA 497
QY 225 ArgValTrpGluGlnAsnThrGlnGlySerIleProTyrValProTrpLeuArgAlaPro 244
Db 498 GACACTCAGGACACAGATGGGCAAGTGGTGTGCCACTGCTCCCTGCTGCAGGGGAG 557
QY 245 TyrTyrValTrpIleGlyArgLeuProSerValGly---HisAlaLeuHisGluGluArg 263
Db 558 CRGCAAGAGCAAGGTGGGCGCTTGGGGAGACTACGATGACAGTGCCTTCATGGAGCCAG 617
QY 264 ValGluArgProProMetPheProPheLeuTyrThrGlnSerTrp-GluAsp-- 282
Db 618 GTACCACGTCCTG-----TGGAGAAGATCT 641
QY 283 -----ProGluProAspMetGluValMe 290
Db 642 GGACAAGCTCCACAGAGCTGCCTGGGGTAAAGTCCCAGAAAGATCTCATGTCAT 701
QY 290 t-----GluIleAsnProLysAsp-----ThrValLeuThrLeuTh 302
Db 702 GCTCAGGACACTGACGCTGAACAAGAGCAAGCAAGAGGAGCTCTCTACATCTGGC 761
QY 302 rSer-----GlyClyCysAsnAlaLeuAsnLeuValGlnGlyAlaGlyGlnValVa 320
Db 762 CTCTGCCAATGGGAATTCAGAACTAGTAAACTCTGCTGGACAGAGCATGTCAACTAA 821
QY 320 lSerValAspCysAsnProAlaGlnSerAlaLeuLeuLeuLysLysValAlaIleG1 340
Db 822 TGTCTCTTGAC--AACAAAAGAGGACAGCTCTGAYAAAG-----GCCGTACA 866
QY 340 nGlnLeuGluPheGluAspValTrpGlnLeuPheGlyGlyGlyValHisProArgIleG1 360
Db 867 ATGCCAGGAAGATGAATGTCGTTAATGTTGCTGGAAACATGGCAGCATGATCCAATATTC 926
QY 360 uGluLeuTyr-GluLysLysLeuAlaProPheLeuSerGlnThrSerHisAsnPheTrpS 380
Db 927 AGATGATATGGAATACCACTCTRCACCTAATGCTTCTTCTAATAATGAAGATAATATG 986
QY 380 erLysArgLeuTrpTyr 385
Db 987 CAAAGCACTGCTCTTAT 1003

RESULT 5
US-11-234-786-369
; Sequence 369, Application US/11234786
; Publication No. US20060024301A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yugu
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
```





QY 283 -----ProGluProaspMetGluValMet-----GluLeuAsnProLysAsp-- 296  
DB 1008 AGTCCCGAGAGAGATCTCATGCTCATGCTCAGGACACCGATGTGAACAAGAGGACAA 1067  
QY 297 -----ThrValLeuThrLeuThrSer-----GlyGlyCysAsnAlaLeuAsnLe 311  
DB 1068 GCAAAAGAGAGACTCTCATCTGCGCTCTGCCAATGGGAATTCAGAAAGTAGTAAACT 1127  
QY 311 uLeuValGlnGlyAlaGlyGlnValValSerValAspCysAsnProAlaGlnSerAlaLe 331  
DB 1128 CGTGTGGACAGACGATGTCACCTTAATGTCCTTGAC---AACAAAGAGAGACAGCTCT 1184  
QY 331 uLeuGluLeuLysValAlaGlnGlnLeuGluPheGluAspValTrpGlnLeuPhe 351  
DB 1185 GACAAAG-----GCCGTACCAATGCCAGGAAGATGTCGCTTAATGTTGCT 1232  
QY 351 eGlyGluGlyValHisProGlnGluGluLeuTyr-GluLysLeuAlaProPheL 371  
DB 1233 GGAACATGGCACTGATCCAAATATTCAGATGATGGAATACCACTCTACACTATGC 1292  
QY 371 euSerGlnThrSerHisAsnPheTrpSerLysArgLeuTrpTyr 385  
DB 1293 TGCTCAATGAAGATAAATTAATGGCCAAAGCACTCTCTAT 1336

## RESULT 8

US-10-932-182A-4215  
; Sequence 4215, Application US/10932182A  
; Publication No. US20060046253A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAO, YOSHIHIRO  
; APPLICANT: NAKAMURA, NORIHISA  
; APPLICANT: KODAMA, YUKIKO  
; APPLICANT: FUJIMURA, TOMOKO  
; APPLICANT: ASHIKARI, TOSHIHIKO  
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS  
; FILE REFERENCE: 030685-043  
; CURRENT APPLICATION NUMBER: US/10/932, 182A  
; CURRENT FILING DATE: 2004-09-02  
; NUMBER OF SEQ ID NOS: 197023  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 4215  
; LENGTH: 3006  
; TYPE: DNA  
; ORGANISM: Saccharomyces pastorianus  
US-10-932-182A-4215

Alignment Scores:  
Pred. No.: 5.92 Length: 3006  
Score: 104.00 Matches: 57  
Percent Similarity: 39.8% Conservative: 43  
Best Local Similarity: 22.7% Mismatches: 79  
Query Match: 3.0% Indels: 72  
DB: 7 Gaps: 15

US-10-620-914-45 (1-648) x US-10-932-182A-4215 (1-3006)

QY 407 AlaValValLeuGlyLeuGlyLysThrValLysArgLeuAlaAsnAlaProThrMetGlu 426  
DB 1672 TCAATTTTAAGAGACTGAATATGATATGATCCCTCAATGGTCAATGTT----- 1719  
QY 427 GluGlnArgLeuTrpAspSerAsnMetLeuLeuHisPheValLysAsnGlyProLys 446  
DB 1720 -----AGTCTCTAGTTAGTGTATTT----- 1740  
QY 447 ProLeuValTrpLeuPheValLysPheValSerLeuValLeu-----PheAsn 462  
DB 1741 -----TGGATCTTTTTCGAATTTTAGGAGTACAGATTTTTCGAAGTTCATTCGA 1791  
QY 463 LysAlaValLeuTrpPheGlyGlyGlyValProGlyLys-----GlnTyrAlaLeuLeu 480  
DB 1792 AGACATGTGTGTGGTATTAATTCGGAAGACCCCTACTGATACATATCAATACAGCCTG---- 1848

QY 481 LysAlaAspGlyIleProIleGluAsnTyrIleAlaArgThrMetAspGlyValAlaGlu 500  
DB 1849 -----CAGTTTGGCGCGCCATCTTGACCCAGTGCACAAA 1884  
QY 501 AsnSerHisValArgLysGlnAsnTyrPheTyrAsnCysLeuThrGlyLys----- 518  
DB 1885 -----CAAGACAAAACTATATCTATGAGGATGGTACTGAAGGTCCTGTTTCA 1932  
QY 519 -----PheLeuArgAspAsnCysProThrTyrLeuArgGluAlaAlaPheAlaThrLeu 536  
DB 1933 AAGGCTTCTT-----TGCCCAACAATATCCCAAGTGTGTTCTTAATGCTAATCCG 1983  
QY 537 LysSerGlyValVal-----AspAsnLeuThrValSerThrAsnPheMetGluGlu 554  
DB 1984 TATAACGGTAGAATTAGCTTTTGATAATATGTCATATCGAATCGAATCTGCTTTGTGCATA 2043  
QY 555 LeuLysAlaArgThrTyrThrLysValIle-----LeuMetAspHisValAspTrpLeu 572  
DB 2044 ATGATGTCTAAACACCTTCACTGATTTGATGATATATACGATGGAT----- 2088  
QY 573 AspMetProValAlaAsnGluLeuAlaGluCysLeuAlaLysGlnValAlaProGlyGly 592  
DB 2089 -----TCAGATGAATGGCTGCATGTTGTTTATCGTTGTTATTTTGTGTC 2136  
QY 593 IleValIleTrpArgSerAlaSerLeuSerProProTyrAlaGluLeuIleGlnLysAla 612  
DB 2137 TTAACCTATTTGG-----CTACTGAATTTACTTATTGTCAGTCTCTGGTG-----TCT 2181  
QY 613 GlyPheAspValArgCysIleArgAlaThrGlnGlyTyrMetAspArgValAsnMet 632  
DB 2182 TCTTTCGAATA-----GCTAACGGAGGTACAGAAGAGAGAACAATA 2226  
QY 633 TyrSerSer-----PheTyrMetAlaArg 640  
DB 2227 TATGTTTCAAAAACATGGGTATATATGCGGT 2259

## RESULT 9

US-10-932-182A-4215  
; Sequence 4215, Application US/10932182A  
; Publication No. US20060046253A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAO, YOSHIHIRO  
; APPLICANT: NAKAMURA, NORIHISA  
; APPLICANT: KODAMA, YUKIKO  
; APPLICANT: FUJIMURA, TOMOKO  
; APPLICANT: ASHIKARI, TOSHIHIKO  
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS  
; FILE REFERENCE: 030685-043  
; CURRENT APPLICATION NUMBER: US/10/932, 182A  
; CURRENT FILING DATE: 2004-09-02  
; NUMBER OF SEQ ID NOS: 197023  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 4215  
; LENGTH: 3006  
; TYPE: DNA  
; ORGANISM: Saccharomyces pastorianus  
US-10-932-182A-4215

Alignment Scores:  
Pred. No.: 5.92 Length: 3006  
Score: 104.00 Matches: 57  
Percent Similarity: 39.8% Conservative: 43  
Best Local Similarity: 22.7% Mismatches: 79  
Query Match: 3.0% Indels: 72  
DB: 7 Gaps: 15

US-10-620-914-45 (1-648) x US-10-932-182A-4215 (1-3006)

QY 407 AlaValValLeuGlyLeuGlyLysThrValLysArgLeuAlaAsnAlaProThrMetGlu 426  
DB 1672 TCAATTTTAAGAGACTGAATATGATATGATCCCTCAATGGTCAATGTT----- 1719  
QY 427 GluGlnArgLeuTrpAspSerAsnMetLeuLeuHisPheValLysAsnGlyProLys 446

```

Db 1720 -----AGTCTATGTTAGTGTATTT-----1740
QY 447 ProLeuValTrpLeuPheValSerLeuValLeu-----PheAsn 462
Db 1741 -----TGGATCTTTTGGATTTTAGGAGTACAGATTTTCAAGGTTCAITTCGA 1791
QY 463 LysAlaValLeuTrpPheGlyGlyValProGlyLys-----GlnTyrAlaLeuLeu 480
Db 1792 AGCAATGTGTGGTATTAATTCGAAGACCTTACTGATACATATCAATACACCGTG---1848
QY 481 LysAlaAspGlyIleProIleGluAsnTyrIleAlaArgThrMetAspGlyLysAlaGlu 500
Db 1849 -----CAGTTTGGCGGGCCATCTTCAGCCAGTGACAAAA 1884
QY 501 AsnSerHisValArgLysGlnAsnTyrPheTyrTyrAsnCysLeuThrGlyLys-----518
Db 1885 -----CAAAAGACAAACTATCTATGAGGATGGTACTGAAGTCTCTGTTTCA 1932
QY 519 -----PheLeuArgAspAsnCysProThrTyrLeuArgGluAlaPheAlaThrLeu 536
Db 1933 AAGGGCTTTCTT-----TGCCCAATATTCGAAGTGTCTTCTAATGCTAATCCG 1983
QY 537 LysSerGlyValVal-----AspAsnLeuThrValSerThrAsnPheMetGluGlu 554
Db 1984 TATAACGGTAGAATTAGCTTTGATATATTTGTCATTCGATGGAACCTTGTCTTTGTCAVA 2043
QY 555 LeuLysAlaArgThrTyrThrLysValIle-----LeuMetAspHisValAspTrpLeu 572
Db 2044 ATGAGTGTCTAACCTTCACTGATTTGATGATTTATATACATGAT-----2088
QY 573 AspMetProValAlaAsnGluLeuAlaGluCysLeuAlaLysGlnValAlaProGlyGly 592
Db 2089 -----TCAGATGAATGGCTGCAATGTTGTTTATCGTTTGTATTTTGTGTC 2136
QY 593 IleValIleTrpArgSerAlaSerLeuSerProProTyrAlaGluLeuLeuGlnLysAla 612
Db 2137 TTAACCTATTGG-----CTACTGAATTACTTATTTCAGTCTCTGGTG-----TCT 2181
QY 613 GlyPheAspValArgCysIleArgAlaThrGlnGlyTyrMetAspArgValAsnMet 632
Db 2182 TCTTTTCAATA-----GCTAACGGAGGTACAGAAGAGAGAAACAATA 2226
QY 633 TyrSerSer-----PheTyrMetAlaArg 640
Db 2227 TATGTTTCAAAAACATGGGTATTATTCGCGT 2259

```

RESULT 10  
 US-10-932-182A-2428  
 ; Sequence 2428, Application US/10932182A  
 ; Publication No. US20060046253A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: NAKAO, YOSHIHIRO  
 ; APPLICANT: NAKAMURA, NORIHISA  
 ; APPLICANT: KODAMA, YUKIKO  
 ; APPLICANT: FUJIMURA, TOMOKO  
 ; APPLICANT: ASHIKARI, TOSHIHIKO  
 ; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS  
 ; FILE REFERENCE: 030685-043  
 ; CURRENT APPLICATION NUMBER: US/10/932,182A  
 ; CURRENT FILING DATE: 2004-09-02  
 ; NUMBER OF SEQ ID NOS: 197023  
 ; SOFTWARE: PatentIn version 3.3  
 ; SEQ ID NO 2428  
 ; LENGTH: 1857  
 ; TYPE: DNA  
 ; ORGANISM: Saccharomyces pastorianus  
 US-10-932-182A-2428

Alignment Scores:  
 Pred. No.: 5.08 Length: 1857  
 Score: 102.00 Matches: 68  
 Percent Similarity: 33.1% Conservative: 36

```

Best Local Similarity: 21.7% Mismatches: 116
Query Match: 2.9% Indels: 94
DB: 7 Gaps: 15

US-10-620-914-45 (1-648) x US-10-932-182A-2428 (1-1857)

QY 84 AsnValAspMetMetAlaAspTyrIleAspLeuAlaLysPheLysSerIleTyrValVal 103
Db 604 AATTGTGAGATGTTCCGTGACTTTACCTGAAGTAAAAAACCAATGAGAAGATAT-----657
QY 104 AspLeuCysHisSerLeuCysGluValAlaLysLysLysAlaLysLysGlyTyrLys 123
Db 658 -----CATTTCTGAAAAACACAAATTCCAAATAATGGTAAATTCGCTTCATGAAA 708
QY 124 AsnValGln-----ValValGluAlaAspAlaCysGlnPheAlaProPro 138
Db 709 GTGGTTCAAGGTTCTCTGTGGATGGCTTTAAGCGCACTAGGGATGAATACCTTCCTCCT 768
QY 139 Glu-----GlyThrAlaThrLeuIleThrPheSerTyrSer 150
Db 769 TCATATGTTTTGGATAGAAAGACTTTCCGGCAAGGCTCTTTATTATTTCAGGTCACATTAC 828
QY 151 LeuThrMetIleProProPheHisAsn-----ValIleAspGln 163
Db 829 TTATTTCTTACTAGGTTTATTTCACAGATTTAAGTATATGCTGCTCGCCATTTCCGAA 888
QY 164 AlaCysSerTyrLeuSerGlnAspGlyLeuValGlyValAlaAspPheTyrValSerGly 183
Db 889 GCATCTGTGTATTTATGTCGATAGGTTAGGTATAATGGT-----924
QY 184 LysTyrAspLeuProLeuArgGlnMetProTyrSerArgArgPheTyrArgSerIle 203
Db 925 ---TATGATCCAAAGACTCAAAAGATCAGATGGGATCGT-----960
QY 204 PheAspIleAspAsnIleAspIleGlyProGluArgAlaTyrLeuGluGlnLysLeu 223
Db 961 -----GTTAGAAATATTGACATTTTCGACCGTAGAACGCAACAAAGTACACGTGAATG 1014
QY 224 GluArgValTrpGluGlnAsnThrGlnGlySerIleProTyrValProTyrLeuArgAla 243
Db 1015 TTGGAGCATGGATATGAACTAAC-----AAGTGGCTGAAGTAC 1056
QY 244 ProTyrTyrValTrpIleGlyArgLeuProSerValGlyHisAlaLeuHisGluGluArg 263
Db 1057 TCCGTTTATTGG-----CGTGTACTAAAAAGGCT-----GAGAAGCCC 1095
QY 264 ValGluArgProProMetPheProThrPheLeuTyrThrGlnSerTrpGluAspPro 283
Db 1096 GGTTCCTCGTTCAACTTTGTTTC-----ACTTCTTAACCTTCAGCATTTTGGCAGGTACC 1149
QY 284 GluProAspMetGluValMetGluIleAsnProLysAspThrValLeuThrLeuThrSer 303
Db 1150 AGACCTGGTTAC-----TACCTGACGTTTGGCCACG 1179
QY 304 GlyGlyCysAsnAlaLeuAsnLeuValGlnGlyAlaGlyGlnValSerValAsp 323
Db 1180 GGGGCG-----TTGTACCAACATGTCGTAAATTTACAGCGCTAAT 1221
QY 324 CysAsnProAlaGlnSerAlaLeuLeuGluLeuLysLysValAlaIleGlnGlnGlu 343
Db 1222 TTCAGGCCA-----ATGTTCTTCGGGAAGATGTTGTAGCCCTCTACCC 1266
QY 344 PheGluAspValTrpGlnLeuPheGlyValHisProArgIleGluGluLeuTyr 363
Db 1267 TCCAAATAATTTATGATCTCATCGGC-----ATATAC 1299
QY 364 GluLysLysLeuAla-----ProPheLeu 371
Db 1300 GCGATCAAGTTAGCATTTGGCTATATGTTACACCACTTCATC 1341

```

RESULT 11  
 US-10-932-182A-2428  
 ; Sequence 2428, Application US/10932182A



```

; Publication No. US20060046253A1
; GENERAL INFORMATION:
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHIRO
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: 030685-043
; CURRENT APPLICATION NUMBER: US/10/932,182A
; CURRENT FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 2428
; LENGTH: 1857
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-10-932-182A-2428

Alignment Scores:
Pred. No.: 5.08 Length: 1857
Score: 102.00 Matches: 68
Percent Similarity: 33.1% Conservative: 36
Best Local Similarity: 21.7% Mismatches: 116
Query Match: 2.3% Indels: 94
DB: 7 Gaps: 15

US-10-620-914-45 (1-648) x US-10-932-182A-2428 (1-1857)

QY 84 AsnValAspMetMetAlaAspTyrIleAspLeuAlaLysPheLysSerIleTyrValVal 103
DB 604 AATTGTGAGATGTTCCGTGACTTACCTGAAAGTAAACCAATGAGAGATAT----- 657
QY 104 AspLeuCyHisSerLeuCyGluValAlaLysLysLysLysLysLysLysLysLysLys 123
DB 658 -----CATTCGGAAAACACAAATTCCAAAAATGTAATGTTGCTTCATGAAA 708
QY 124 AsnValGln-----ValValGluAlaAspAlaCysGlnPheAlaProPro 138
DB 709 GTGGTTCAAGGTTCTCTGTGGATGCTTTAAGCGCACTAGGGATGAAATACCTCCCTCT 768
QY 139 Glu-----GlyThrAlaThrLeuIleThrPheSerTyrSer 150
DB 769 TCATATGTTTTGGATGAAAGACTTTCGGGCAAGGTCCTTTATTTTCAGGTCACATTAC 828
QY 151 LeuThrMetIleProProPheHisAsn-----ValIleAspGln 163
DB 829 TTATTTCTTAGTGTGTTTATTCAGATTTAAGTATTATGCTGCTGGACCATTTCCGAA 888
QY 164 AlaCysSerTyrLeuSerGlnAspGlyLeuValGlyValAlaAspPheTyrValSerGly 183
DB 889 GCATCTGTGTTATGTTGATTTAGGTATATATGTT----- 924
QY 184 LysTyrAspLeuProLeuArgGlnMetProTyrSerArgArgPheThrArgSerIle 203
DB 925 ---TATGATCAAGACTCAAAAGATCAGATGGGATCGT----- 960
QY 204 PheAspIleAspAsnIleAspIleGlyProGluArgAlaTyrLeuGluGlnLysLeu 223
DB 961 -----GTAGAAATATGATTTGGACCGGTAGAACGGCAAAAGTACACGTGAATG 1014
QY 224 GluArgValTrpGluGlnAsnThrGlnGlySerIleProTyrValProTyrLeuArgAla 243
DB 1015 TTGGAGCATGGAATATGACACTAAC-----AAGTGGCTGAAGTAC 1056
QY 244 ProTyrTyrValTrpIleGlyArgLeuProSerValGlyHisAlaLeuHisGluGluArg 263
DB 1057 TCCGTTTATTTG-----CGTGTACTAAAAAGGGT-----GAGAAGCCC 1095
QY 264 ValGluArgProMetPheProThrPheLeuTyrThrGlnSerTrpGluAspPro 283
DB 1096 GGTTCCTGCTCAACTTTGTTTC-----ACTTCTTAACATTTTGGCAGCGTACC 1149

```

---

```

QY 284 GluProAspMetGluValMetGluIleAsnProLysAspThrValLeuThrLeuThrSer 303
DB 1150 AGACCTGGTTAC-----TACCTGACGTTTGGCAGC 1179
QY 304 GlyGlyCyAsnAlaLeuAsnLeuValGlnGlyAlaGlyGlnValValSerValAsp 323
DB 1180 GGGGCG-----TTGTACCAAAACATGTGGTAAATTTACAGGCGTAAAT 1221
QY 324 CysAsnProAlaGlnSerAlaLeuLeuGluLeuLysValAlaIleGlnGlnLeuGlu 343
DB 1222 TTCAGGCCA-----ATGTTCTTACGGGAAGATGTTGCTTACCGCTCTACCC 1266
QY 344 PheGluAspValTrpGlnLeuPheGlyGlyGlyValHisProArgIleGluLeuTyr 363
DB 1267 TCCAAAATAATTTATGATCTCATCGGC-----ATATAC 1299
QY 364 GluLysLysLeuAla-----ProPheLeu 371
DB 1300 GCGATCAAGTTAGCATTTGGCTATATGTTACCAACCATTTATC 1341

RESULT 12
US-11-124-367A-21
; Sequence 21, Application US/11124367A
; Publication No. US20060024700A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: Hongjin Huang
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; FILE REFERENCE: CL001519.ORD
; CURRENT APPLICATION NUMBER: US/11/124,367A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,846
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/582,609
; PRIOR FILING DATE: 2004-06-25
; PRIOR APPLICATION NUMBER: US 60/599,554
; PRIOR FILING DATE: 2004-08-09
; NUMBER OF SEQ ID NOS: 34460
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 3042
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-11-124-367A-21

Alignment Scores:
Pred. No.: 9.75 Length: 3042
Score: 102.00 Matches: 90
Percent Similarity: 33.3% Conservative: 54
Best Local Similarity: 20.8% Mismatches: 163
Query Match: 2.9% Indels: 126
DB: 12 Gaps: 19

US-10-620-914-45 (1-648) x US-11-124-367A-21 (1-3042)

QY 12 TyrThrLysLysAsnPheSerLeuGluLysLeuLysSerMetLysAspAspLeu 31
DB 199 TTTTACCGGACGTAGTTTCAGATCGTGAATTCNAA---GCCACATGTGCAACCTACTG 255
QY 32 ThrValLeuArgHisMetTrpPheGlySerLysLysGlyAspAspHisAlaAla----- 49
DB 256 GCGTATCTAAGCACCTC-----AAAGGGCAAAACCGAGCGCCCTGGAA 300
QY 50 -----ArgLeuGluSerPheTyrGlyProGlnAlaAla-----AlaPheAlaAla 64
DB 301 TCGTTACGTAAAGCTGAAGAGTAAATCCAGCAAGAGCATGCTGACCGAGCAAGAAATCAGA 360
QY 65 ArgLeuAlaGluArgSerAsnLeuIleTrpValAspLeuGlyGlyThrGlyGluAsn 84
DB 361 AGTCTGGTCACCTGGGAAACTATGCTGGGTCTACTATCACAATGCGGCGGCTCTCAGAC 420
QY 85 ValAspMetMetAlaAspTyrIleAsp-----LeuAlaLysPheLysSerIleTyrVal 102

```



195 SerArgArgPhePheTrpArgSerIlePheAspIleAspAsnIleAspIleGlyProGlu 214  
4219 GAAAGAACTAGAAAGTGGGAGATGCACCTA----- 4248  
215 ArgArgAlaTyrLeuGluGlnIleValTrpGluGlnAsnThrGlnGlySer 234  
4249 ---CACGCTTATTAATGAAGT-----GAGAGGAGGAGAGATACATCT 4287  
235 IleProTyrValProTrpLeuArgAlaProTyrTyrValTrpIleGlyArgLeuProSer 254  
4288 GTAGGT-----GTTACTCTTGGTAGATGAGATCT 4317  
255 ValGlyHisAlaLeuHisGlu-----GluArgValGluArg 266  
4318 CTT---CATGCCCTAGTGAATGGAGCAATATCTCAACTAGCTCTAGAAAATGGAGA 4374  
267 ProProMetPheProProThrPheLeuTyrThrGln----- 278  
4375 ATATCTAAATATCCGACGAGAAATTAATTTGCTCTTTGGCAGCGGTGCAGATGGGT 4434  
279 -----SerTrpGluAspProGluProAspMetGluValMetGluIleAsnProLys--- 295  
4435 TTAGGCGAATGGATATGCTTGGAGCAATATATTAGTGTATGAAACCAATTTCCCGAT 4494  
296 -----AspThrValLeuThrLeuThrSerGlyGly----- 305  
4495 AAGGAATTTCTTGTATGATCAATATATATTATGACAAAAACGATPACGAGGATGCTGTATA 4554  
306 -----CysAsnAlaLeuAsnLeuValGlnGlnGlyAlaGlyGlnValValSerValasp 323  
4555 CACATCTAAATGCCAGAGATTTGCTGGTAACCTGAAATATCCGCATCAATTAACGAAAGT 4614  
324 CysAsnProAlaGlnSerAlaLeuLeuGluLeuLysValAlaIleGlnGlnLeuGlu 343  
4615 TACATAGAGCTTATAGTGAATTTGTAGAACACAAATATATCA-----GAG 4662  
344 PheGluAspValTrpGlnLeuPheGlyGluGlyValHisProArgIleGluGluLeuTyr 363  
4663 TTGAAGAAATA-----ATTAAA 4680  
364 GluLysLeuAlaProPheLeuSerGlnThrSerHis-----AsnPheTrpSerLys 381  
4681 TATAAGCAACTGGCGCTTAACCGAGAGAAATTTCAATATCAAAACCTTTGGACGAAA 4740  
382 ArgLeuTrpTyrPheGlnHisGlyLeuTyrTyrGlnGlyMetGlyLysLeuCysTrp 401  
4741 AGGTTG----- 4746  
402 ValLeuGlnCysLeuAlaValValLeuGlyLeuGlyLysThrValLysArgLeuAlaAsn 421  
4747 -----CTGGTTGCCAAAAGAACGTTGAT----- 4770  
422 AlaProThrMetGluGlnArgArgLeuTrpAspSerAsnMetLeuIleHisPheVal 441  
4771 -----TTATGGCAAGAGATCACTAGAGTGAATCATTA 4803  
442 LysAsnGlyProLysProLeuValTrpLeuPheValLysPheValSerLeuVal----- 459  
4804 GTGATTAACCAACCAAGATTTACAATATGGATCAAGTTTCTTAATCTGTGTAGAAAA 4863  
460 -----LeuPheAsnLysAla-----ValLeuTrpPheGlyGlyVal 472  
4864 TCCGGTAGAATGAGATGGCTCAAAAGGCACTGAATATGCTTCTTCAAGGAGCGGTGAT 4923  
473 Pro 473  
4924 CCT 4926

## RESULT 14

US-10-932-182A-1448

; Sequence 1448, Application US/10932182A

; Publication No. US20060046253A1

GENERAL INFORMATION:  
; APPLICANT: NAKAO, YOSHIHIRO  
; APPLICANT: NAKAMURA, NORIHIRO  
; APPLICANT: KODAMA, YUKIO  
; APPLICANT: FUJIMURA, TOMOKO  
; APPLICANT: ASHIKARI, TOSHIHIKO  
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS  
; FILE REFERENCE: 030685-043  
; CURRENT APPLICATION NUMBER: US/10/932,182A  
; CURRENT FILING DATE: 2004-09-02  
; NUMBER OF SEQ ID NOS: 197023  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 1448  
; LENGTH: 7437  
; TYPE: DNA  
; ORGANISM: Saccharomyces pastorianus  
US-10-932-182A-1448

Alignment Scores:  
Pred. No.: 35.9 Length: 7437  
Score: 101.50 Matches: 102  
Percent Similarity: 32.0% Conservative: 71  
Best Local Similarity: 18.9% Mismatches: 147  
Query Match: 2.9% Indels: 221  
DB: 7 Gaps: 27

US-10-620-914-45 (1-648) x US-10-932-182A-1448 (1-7437)

QY 32 ThrValLeuArgHisMetTrpPheGlySer----- 41  
DB 3670 TCAGTTTTGAAGAGCGCTTGGAAATTCACGCCAACAAAGAACTAAAGAGGACTGGCAAGAA 3729  
QY 42 -----LysLysGlyAspAspHisAlaAlaArg--- 50  
DB 3730 TGGAGTAAGGATTATCTATTCAATTGCTTAAGAAATCCCTTCCACGCACTGAGAGCT 3789  
QY 51 -----LeuGluSerPheTyrGlyProGlnAla-----AlaAlaPhe 62  
DB 3790 TGTTCGAATCTCGCTAGTAGTACTATCGGTAGCTAGAGAACTTTCAATACTGCTTTC 3849  
QY 63 AlaAlaArgLeuAlaGluArgSerAsnLeuIleTrpValAspLeuGlyGlyGlyThrGly 82  
DB 3850 GCAAGT-----ATTGGACGGAGCTTTATAGCCAGTATCAA 3885  
QY 83 GluAsnVal-----AspMet 87  
DB 3886 GAAGATTGATCGAGTCGCTATGCATAGCTTCTCTCTCAAAATCCGCCGGAATA 3945  
QY 88 MetAlaAspTyrIleAspLeuAlaLysPhe-----LysSerIleTyrVal 102  
DB 3946 CATCAAAACGTTATAAATTTGGTTCGAATTCATGGAGCATGATGAAGCAATTCCTATC 4005  
QY 103 -----ValAspLeuCysHisSerLeuCysGluValAlaLys 114  
DB 4006 CCAACTCAAAATTTGGTGAATATGCTGAGAGATGCCAGCATAT----- 4050  
QY 115 LysLysAlaLysAlaLysGlyTrpLysAsnValGlnValAlaGluAlaAspAlaCysGln 134  
DB 4051 -----GCTAAGCATACATTACAAGAAATTAATTTATCAAA----- 4089  
QY 135 PheAlaProGluGlyThrAlaThrLeuIleThrPheSerTyrSerLeuThrMetIle 154  
DB 4090 -----GAGCCTGAAAACCTCC-----ACAATTTGAATCACTT 4119  
QY 155 ProProPheHisAsnValIleAspGlnAlaCysSerTyrLeuSerGlnAspGlyLeuVal 174  
DB 4120 ATCAGTATCAATAATCAATTGAACCAACG-----GACGCTGCAATT 4161  
QY 175 GlyValAlaAspPheTyrValSerGlyLysTyrAspLeuProLeuArgGlnMetProTrp 194  
DB 4162 GGTATTTTAAAG---CACGCCAGCAGCAACCATTCGTTGCAATTAAGGAAACATGTTTC 4218  
QY 195 SerArgArgPheTrpArgSerIlePheAspIleAspAsnIleAspIleGlyProGlu 214

Db 4219 GAAAACTAGAAAAGTGGGAAGATGCACTA----- 4248  
Qy 215 ArgArgAlaTyrLeuGluGlnLysLeuGluArgValTrpGluGlnAsnThrGlnGlySer 234  
Db 4249 ---CACGCTTATAATGAACGT-----GAGAAGCGCAGAGATACATCT 4287  
Qy 235 IleProTyrValProTyrLeuArgAlaProTyrTyrValTrpIleGlyArgLeuProSer 254  
Db 4288 GTAGGT-----GTTACTCTTGTAAGATGAGATCT 4317  
Qy 255 ValGlyHisAlaLeuHisGlu-----GluArgValGluArg 266  
Db 4318 CTT---CATGCCCTAGTGAATGGAGCAATTATCTCAACTAGCTGCTAGAAAATGGAGA 4374  
Qy 267 ProProMetPheProThrPheLeuTyrThrGln----- 278  
Db 4375 ATATCTAAATTTACCGACGAAAGAAATTAATTTCTCTTTGGCAGCGGTGCAGCATGGGT 4434  
Qy 279 -----SerTrpGluAspProGluProAspMetGluValMetGluIleAsnProLys--- 295  
Db 4435 TTAGCGAATGGATATGCTTGGCAATATATTAGTGTATGTAATGAACCAAAATTTCCCGAGAT 4494  
Qy 296 -----AspThrValLeuThrLeuThrSerGlyGly----- 305  
Db 4495 AAGGAATTTCTTGTGCAATATTATATTTCACAAAAACGATTACGAGGATGCTGGTAAA 4554  
Qy 306 -----CysAsnAlaLeuAsnLeuLeuValGlnGlyAlaGlyGlnValValSerValAsp 323  
Db 4555 CACATACTAAATGCCAGAGATTTGCTGGTAATATCCGCACTAATTAACGAAAGT 4614  
Qy 324 CysAsnProAlaGlnSerAlaLeuLeuGluLeuLysLysValAlaIleGlnLeuGlu 343  
Db 4615 TACATAGACTTATAGTGTAAATGTTAGAACACAAATAATACAA-----GAG 4662  
Qy 344 PheGluAspValTrpGlnLeuPheGlyGluGlyValHisProArgIleGluGluLeuTyr 363  
Db 4663 TTGAAGAAATA-----ATTAAA 4680  
Qy 364 GluLysLysLeuAlaProPheLeuSerGlnThrSerHis-----AsnPheTrpSerLys 381  
Db 4681 TATAAGCAACTGGCGCTAACTCGGAGAAAAAACTTCATTTATCAAAACCTTTGGACGAAA 4740  
Qy 382 ArgLeuTrpTyrPheGlnHisGlyLeuTyrTyrGlnGlyMetGlyLysLeuCysTrp 401  
Db 4741 AGGTTG----- 4746  
Qy 402 ValLeuGlnCysLeuAlaValValLeuGlyLeuGlyLysThrValLysArgLeuAlaAsn 421  
Db 4747 -----CTGGGTTGCCAAAGAACGTTGAT----- 4770  
Qy 422 AlaProThrMetGluGluGlnArgArgLeuTrpAspSerAsnMetLeuIleHisPheVal 441  
Db 4771 -----TTATGGCAAAAGAGTACTCTCAGAGTGAGATCATTA 4803  
Qy 442 LysAsnGlyProLysProLeuValTrpLeuPheValLysPheValSerLeuVal----- 459  
Db 4804 GTGATTAACCAAAACAAGATTACAAATATGATCAAGTTTGTCTAATCTGTGTAGAAA 4863  
Qy 460 -----LeuPheAsnLysAla-----ValLeuTrpPheGlyGlyVal 472  
Db 4864 TCCGGTAGAATGAGACTGGCTCAAAAGGCACTGAATATGCTTCTTGAAGGAGCGGTGAT 4923  
Qy 473 Pro 473  
Db 4924 CCT 4926

RESULT 15  
US-11-169-041-29  
; Sequence 29, Application US/11169041  
; Publication No. US20060019284A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company

; TITLE OF INVENTION: IDENTIFICATION OF POLYNUCLEOTIDES FOR PREDICTING ACTIVITY OF  
; TITLE OF INVENTION: COMPOUNDS THAT INTERACT WITH AND/OR MODULATE PROTEIN TYROSINE  
; TITLE OF INVENTION: KINASES AND/OR PROTEIN TYROSINE KINASE PATHWAYS IN LUNG CANCER  
; TITLE OF INVENTION: CELLS  
; FILE REFERENCE: 10001 NP  
; CURRENT APPLICATION NUMBER: US/11/169,041  
; PRIOR FILING DATE: 2005-06-28  
; PRIOR APPLICATION NUMBER: 60/584,405  
; PRIOR FILING DATE: 2004-06-30  
; NUMBER OF SEQ ID NOS: 527  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 29  
; LENGTH: 3063  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; US-11-169-041-29  
  
Alignment Scores: 12.5 Length: 3063  
Pred. No.: 101.00 Matches: 90  
Score: 33.3% Conservative: 54  
Percent Similarity: 20.8% Mismatches: 163  
Best Local Similarity: 2.9% Indels: 126  
Query Match: 12 Gaps: 19  
DB: 19  
  
US-10-620-914-45 (1-648) x US-11-169-041-29 (1-3063)  
  
Qy 12 TyrThrLysLysAsnPheSerLeuGluLysLeuLysSerSerMetLysAspAspLeu 31  
Db 198 TTTTACCGGACTGAGTTCAGATTCGTAATCTCAAA---GCCACAATGTCACACTACTG 254  
Qy 32 ThrValLeuArgHisMetTrpPheGlySerLysLysGlyAspAspHisAlaAla----- 49  
Db 255 GCCTATCTAAAGCACCTC-----AAAGGCAAAACGAGGCGAGCTGGAA 299  
Qy 50 -----ArgLeuGluSerPheTyrGlyProGlnAlaAla-----AlaPheAlaAla 64  
Db 300 TGCTTACGTAAGACTGAAGAGTTAATCCAGCAAGAGCATGCTGACCAGGCGAGAAATCAGA 359  
Qy 65 ArgLeuAlaGluArgSerHisLeuIleTrpValAspLeuGlyGlyThrGlyGluAsn 84  
Db 360 AGTCTGGTCACCTGGGGAACATATGCTGGGTCTACTATCACATGGGCCGACTCTCAGAC 419  
Qy 85 ValAspMetMetAlaAspTyrIleAsp-----LeuAlaLysPheLysSerIleTyrVal 102  
Db 420 GTTCAGATTATGTAGCAAGGTGAGACATGCTGTGAGAAGTTTTCAGTCCCTATAGA 479  
Qy 103 ValAspLeuCysHisSerLeuCysGluValAlaLysLysLysAlaLysGlyTrp 122  
Db 480 ATTGAGAGTCCAGAGCTTGAAGAGGCTGAGAGGCTGAGACACGCTTAAAGTGTGGAGGAAC 539  
Qy 123 LysAsn-----ValGlnValValGluAlaAspAlaCysGlnPheAlaPro-----Pro 138  
Db 540 CAAAATGAAGAGCAAGAGGTGCTCTTTGAGAAGGCTCTGGAAGAGCAAGCAAGCAACCA 599  
Qy 139 GlnGlyThrAlaThrLeuIleThrPheSerTyrSerLeuThrMetIleProProPheHis 158  
Db 600 GAATTCACCTCTGGATGGCAATAGCAAGTACCGTCTGGCACTGGCCACCATCTCAG 659  
Qy 159 AsnValIleAspGlnAlaCysSerTyrLeuSerGlnAspGlyLeuValGlyValAlaAsp 178  
Db 660 AAGCCATTGAC-----CCTCTGAGGCAAGCC----- 686  
Qy 179 PheTyrValSerGlyLysTyrAspLeuProLeuArgGlnMetProTrpSerArgArgPhe 198  
Db 672 -----CCTCTGAGGCAAGCC----- 686  
Qy 199 PheTrpArgSerIlePheAspIleAspAsnIleAspIleGlyProGluArg----- 215  
Db 687 -----ATTCCGCTGAATCTCTGACACCACTACCTT 716  
Qy 216 ArgAlaTyrLeuGluGlnLysLeuGluArgValTrpGlnGlnAsnThrGlnGlySerIle 235

```
Db 717 AAGTCTCTCGGCTCGAAGCTTCATAAGATCGGTGAAGAA----- 758
Qy 236 ProTyrValProTyrLeuArgAlaProTyrTyrValTyrIleGlyArgLeuProSerVal 255
Db 759 -----GGTGAAGAGGAAGGTGAA 776
Qy 256 GlyHisAlaLeuHisGluGluArgValGluArgProPro-----MetPhePro 271
Db 777 GGAGAGAAGTTAGTTGAAGAAGCTTTGGAGAAAGCCCGAGGTGTAAACAGATGTACTTCGC 836
Qy 272 ProThrPheLeuTyrThrGlnSerTyrGluAspProGluProAspMetGluVal----- 289
Db 837 AGTGCAGCCCAAGTTTATCGAAGAAAGATGACCCAGCAAGCGATTGAACCTGCTTAA 896
Qy 290 -----MetGluIleAsnProLysAspThrValLeuThrLeuThrSerGlyGlyCysAsn 307
Db 897 AAGGCTTTAGAAATACATACCAACAATGCCCTACTGCAATGCCCAATTTGGTGCTGTAT 956
Qy 308 AlaLeuAsnLeuValGlnGlyAlaGlyGlnValValSerValAspCysAsnPro--- 326
Db 957 AGGGCAAAAGTCCTC-----CAAGTAATGAATCTAAGAGAGAATGGAATG 1001
Qy 327 ---AlaGlnSerAlaLeuLeuGluLeuLysValAlaIleGlnGlnLeuGluPheGlu 345
Db 1002 TATGGGAAAGAAAGTTACTGGAACTAATAGGACACGCTGTGGCTCATCTGAAGAAAGCT 1061
Qy 346 Asp-----ValTyrGlnLeuPheGly----- 352
Db 1062 GATGAGGCCAATGATATCTCTCCGTGTCTGTTCATCTTCCAGCCTCCATGCTCTA 1121
Qy 353 -----GluGlyValHisProArgIleGluGluLeuTyrGluLysLysLeuAla 368
Db 1122 GCAGATCAGTATGAAGAAGCAGAGTATTACTTCCAAAAGGAATTCAGTAAAGAGCTTACT 1181
Qy 369 ProPheLeuSerGlnThrSerHis-----AsnPheTrp----- 379
Db 1182 CCTGTAGCGAAACAACCTGCTCCATCTGCGGTATGGCAACTTTCAGCTGTACCAATGAAG 1241
Qy 380 -----SerLysArgLeuTyrPheGlnHisGlyLeu 390
Db 1242 TGTGAAGACACAGGCCATCCACCACTTTATAGAGGGTGTA 1280
```

Search completed: March 14, 2006, 07:27:55  
Job time : 1251 secs

**THIS PAGE BLANK (USPTO)**